A Comparison of Typically Developing and Atypically Developing ToM

Yasmin Harman-Smith, BHSc (Psych Hons), BA

School of Psychology

University of Adelaide

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

November, 2009

Table of Contents

Summary

This thesis examines Theory of Mind (ToM) in relation to the areas of language and cognitive development. The thesis explores both popular and alternate theories of ToM and how they account for the important relationships between language and theory of mind. It examines the theories in the context of published ToM findings as well as the findings from three studies conducted by the author.

The first study took the form of a pilot study which re-analysed data, collected for the author‟s honours project, from a small group of children with (n = 10) and without autism (n = 10). In each diagnostic group (autism and no autism) children were divided into two groups, those passing a ToM task and those failing a ToM task. The aim of the study was to investigate whether the underlying language and cognitive skills required to succeed on ToM tasks are the same for children with and without autism. The key finding of the study was that for both the children with and without autism, those who passed the ToM task performed better on all the developmental measures, although only the difference in language ability was statistically significant.

The second study expanded on the pilot study, examining the relationship in typically developing 4 and 6-year-old children, between ToM, language, cognitive development, and subtractive reasoning. The study‟s aim was to examine the developmental structure underlying ToM using factor analysis. The results indicated that for 4-year-old children the most important skill for ToM success was language, but that for 6-year-old children ToM success was more strongly related to subtractive reasoning ability. The findings of the study also raised the question of whether presentation method for ToM tasks impacted on task difficulty.

A final study therefore examined the effect of presentation mode 2-dimensional versus 3-dimensional, on the success of typically developing 4-year-old children on the ToM task battery.

The findings indicated that tasks presented in cartoon format were more difficult than tasks presented with dolls and props.

Reliability and validity of common ToM tasks and new ToM test batteries are discussed. Alternative conceptions of ToM in relation to social interaction are considered.

Declaration

I declare that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. I also give permission for the digital version of my thesis to be made available on the web, via the University"s digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Yasmin Harman-Smith

Signed ____ Date ____________________

Acknowledgements

I am forever grateful for the guidance and support provided throughout my PhD candidature. My principal supervisor, Ted Nettelbeck, shared his wisdom and expertise at every opportunity. I acknowledge that without the benefit of his insight and knowledge, the PhD process would have been far more treacherous and troublesome. I am indebted to Ted for his support, his commitment and dedication to reading drafts promptly, and his belief in me, all of which have lessened the burden of the PhD journey.

My secondary supervisor, Lisa Kettler, also provided valuable advice and guidance over the past 3 and a half years. I always left „chats" with Lisa, whether scheduled or in passing, feeling rejuvenated and clear headed. Lisa"s feedback on my work was always helpful in fine tuning explanations and untangling complex ideas. Lisa"s practical advice and compassion helped me navigate through the minefield that became life when PhD met motherhood.

Too many other colleagues to name have helped shape my ideas, understanding, and work ethic. Particularly, I would like to thank all the wonderful women who I had the pleasure of working alongside at the Flinders Early Intervention Research Program. I would also like to thank my colleagues in the School of Psychology, the teaching staff, professional staff, and my fellow PhD students. Moving on from here, I will miss the camaraderie within the school. People were always willing to chat and debrief, mind a baby, read drafts, provide statistical support, and the list goes on.

My sincerest thanks go to the children, families, teachers, childcare centres, preschools, and schools who participated in my research and were always cheerful despite the interruption to their busy lives.

I also extend my gratitude to the Department of Education and Child Services (DECS) for assisting with the process of recruiting participants and for sharing with me their ideas on my research.

Emotional and practical support was provided by dear family and friends. Each helped in their own way, and to each I am indebted. Thank you Tony, Alexis, Elise, Elle, Gran, Marnie, Colleen, Brian, Sue, Peter, Travis, Barbara, Joanne H., Kate, Carolyn, Josie, Luke, Julia, James, Helen, Tod, Fiona, Joanne C., Tess, Ian, Ryan, Shona, and Vic. I must especially thank my mother, Elisabeth, who has become a second mother to my daughter, and without whom completing this thesis would not have been possible.

Lastly, I would like to thank my darling husband and delightful daughter. Mark and Emelia grounded me in the real world during turbulence and chaos. Mark‟s love and support was and continues to be unconditional. In addition to being a wonderful husband, I must also thank him for our amazing daughter. Seeing Emelia at the end of each day, aside from helping put life in perspective, was the ultimate reward for a long day‟s work. Coming home to her smile, loving cuddles and kisses, and "I missed you Mummy, how was your day?" made every day, no matter how challenging, wonderful.

PREFACE - Synthesis of the Studies Comprising this Thesis

This thesis is comprised of both traditional chapters and manuscripts. Specifically, Chapters 3, 4, and 6 of this thesis are presented in manuscript format. These manuscripts have been prepared for publication, but at this time have not been published. The first manuscript, Chapter 3, was presented at the Biennial National Autism Conference, Gold Coast, Qld, Australia, March 2007. It is anticipated that Chapters 4 and 6 will be submitted for review either separately or as a combined manuscript. As a result of the manuscript format of the aforementioned chapters some redundancy in the introduction sections of the manuscripts is inevitable, because to some extent these introductions reflect the literature reviewed in preceding chapters. It is also the case that all references cited within each of the manuscript chapters have been collated and included in the single reference section, beginning on page 165. With the aim of synthesising the thesis a brief overview of the manuscripts and rationale for each study is presented here.

Pilot Study - Chapter 3

The pilot study, in which I reanalysed the data collected from my honours project, was conceived to examine the viability of a larger study designed to compare abilities underlying ToM in children with autism with abilities underlying ToM in typically developing children. The main idea for the honours study stemmed from personal experience in working closely with children with autism in an early intervention setting. I was fortunate to be able to observe the learning of children with autism across a number of years while conducting one-on-one intervention. It occurred to me that when children with autism were engaged in less complicated social interaction - goal directed interaction that limited excessive communicative cues - children were better able to focus on the communicative exchange. Thus, the honours study was devised to examine whether early one-on-one intervention improved social interaction skills in children with autism as a by-product of the focused communicative exchanges. ToM was chosen as a measure of social ability because of its prominent role in early social developmental research, but also because ToM was thought to be easily assessed with an experimental task, the false-belief test.

Although the honours sample was limited, it appeared that ToM ability in children with autism was relatively good in comparison to what was expected in light of findings presented in the literature. Not only did children with intervention perform well, but those without intensive intervention had also performed better than expected. For this reason it appeared to be important to examine the factors that might have contributed to the seemingly above average ToM performance. Thus the pilot study, presented in Chapter 3, was devised. The rationale for the reanalyses was that comparing developmental abilities in relation to ToM ability might better explain ToM success in autism relative to ToM success in typically developing children. The literature reviewed in Chapters 1 and 2 highlighted that there were some similarities but also some differences in the ways in which ToM was related to early child development in autism when compared to typical development. Generally the literature used a wide range of measures to assess developmental aspects of interest, and often these measures were different for different populations of children. For that reason it seemed sensible to examine differences in skills assessed by a single measure that was appropriate for use with both children with autism and children without autism. The honours data were deemed to be acceptable for a pilot analysis because all children had been assessed with the same developmental measure and the same ToM task.

Study 1 - Chapter 4

Following analysis of the pilot data my supervisors and I discussed the design for subsequent research. The aim of future research was to examine in more detail the differences and similarities in skills that the pilot analysis had found to be related to ToM in children with autism and normally developing children. During these discussions my supervisors pointed out that intelligence research showed that in young pre-school children relatively independent cognitive abilities that subsequently emerge are not yet well differentiated. We considered that this could contribute to findings when examining relationships between variables. In response to this possibility I decided to incorporate two well defined age groups in order to examine how ToM relates to development when ToM begins to emerge but also later in childhood when children"s abilities are better differentiated from one another. Thus, the study presented in Chapter 5 was conceived.

Study 2 - Chapter 6

Study 2 expanded on questions raised by results from Study 1. The test battery developed for use in Study 1 appeared to be inordinately difficult for both age groups participating in the study. One strong possibility was that this might have been a result of the materials used to present ToM tasks. For ease of presentation and to ensure presentation consistency, the ToM tasks in Study 1 had been presented with cartoon strips and pictures, rather than dolls and props. To assess the plausibility of this explanation, an additional review of the literature was conducted (Chapter 5 of this thesis), specifically in relation to reliability and validity of ToM tasks. Although much of this literature had already been considered, it seemed appropriate to consider this literature in closer detail. Closer examination of this literature found no studies in which task difficulty as a result of presentation method had been experimentally investigated. Hence, the second study was conceived to address this gap in the literature. Principally, the study aimed to address whether method of presentation could impact task difficulty sufficiently to account for the perplexing difficulty of the task battery used in Study 1.

In summary, the studies comprising this thesis address questions raised both in the literature and questions arising from studies conducted previously by the current author.

CHAPTER 1 - Theory of Mind: Background and Theory

Introduction to Theory of Mind

This thesis examines Theory of Mind (ToM) in relation to early child development in the areas of language and reasoning. The literature has treated the term ToM in three main ways. Firstly, the phrase „ToM research‟ has been used to describe a broad area of research which investigates how people think about thinking. Secondly, ToM is a term that has been applied to the human capacity to understand that other people have thoughts, beliefs, and desires. Thirdly, the term ToM has also been used to describe what is measured when research investigates people‟s ability to assess what these thoughts, beliefs, and desires might be. This thesis is concerned mainly with how ToM is measured, that is how the ability to think about thinking is assessed in human populations. This chapter presents the main developmental theories of ToM in order to provide a foundation for discussing the measurement of ToM and its developmental correlates. Theories are discussed in relation to both experimental and non-experimental findings from the literature.

In order to establish a framework for ToM, a brief account of the beginnings of ToM research follows. While philosophising about the nature of the human mind and interaction with the mind dates back as far as written records (MacDonald, 2003), the research that spurred the debate most relevant to this thesis began in the 1970s. This research investigated chimpanzees‟ abilities to predict the behaviour of other chimps (Premack & Woodruff, 1978). Philosophers were interested in whether a chimp could predict the behaviour of another chimp, based on a concept held in the mind of the first chimp about what the other chimp appeared to believe. This area of philosophising lead to the conception of the False-belief test for use with children, which assesses an individual‟s ability to predict the behaviour of another person based on that person‟s beliefs (Wimmer & Perner, 1983). In other words, this early research investigated people‟s mental explanations of behaviour; specifically the age at which children begin to understand that people‟s behaviour can be attributed to mental states. The tests devised to measure children‟s ability to attribute mental states were termed tests of false-belief.

Tests of false-belief assess the ability to infer the behaviour of a person by considering what the other person knows. Successfully completing a false-belief test involves, firstly inferring what another person can be expected to know, given a series of events; secondly, appreciating that beliefs govern behaviour; and thirdly, reasoning that a given belief will result in a logical behaviour. An example of a false-belief test, also known as the Sally-Anne task (an unexpected transfer task) involves two characters, Sally and Anne. As described by (Frith & Frith, 1999), a series of events leads to Sally having a misrepresentation of reality. The story, illustrated with pictures or dolls, is as follows: Sally has a ball and she puts it in a basket. Then Sally leaves the scene. After which, Anne moves the ball to a box. Finally, Sally returns to the scene and wants to retrieve her ball. At this point the child is asked "Where will Sally look for her ball?". The test question targets the understanding that Sally has a false-belief. A belief which she has because she did not see the ball being moved; she was not present at the time. In other words, Sally‟s mental representation (belief) of the location of the ball no longer matches the ball‟s actual location. Therefore, the precursor to successfully reasoning about Sally‟s action when she returns and wants to get her ball is firstly correctly representing Sally‟s representation; referred to as meta-representation (i.e., thinking about thinking). The second requirement is an understanding that Sally‟s belief will influence her behaviour. The final requirement is successfully combining these competencies. That is, reasoning correctly:

> Series of events → Sally thinks ball is in basket (belief), Sally thinks ball is in basket (belief) + Sally wants ball (desire) → Sally looks in basket (action).

Since Wimmer and Perner‟s inaugural study of ToM in children, a substantial body of research, some of which was summarised in a meta-analysis (Wellman, Cross, & Watson, 2001), has established that an ability to answer these types of false-belief questions begins to emerge in children around the age of 4 years. Furthermore, children whose early social interactions are disrupted in some way, experience delays or deficits in ToM; for instance children with autism who do not orient to social stimuli (Baron-Cohen, Leslie, & Frith, 1985) and children who are born deaf to hearing parents who receive limited language input (Peterson & Siegal, 1995). Standard false-belief tests have, however, been criticised for being unduly difficult for young children. It has been argued that non-ToM related task artefacts, such as language complexity and cognitive processing load, obscure earlier ToM understanding (Bloom & German, 2000). Bloom and German argued that children younger than 4 years are able to pass tests of ToM that are less cognitively complex than the standard Sally-Anne type false-belief task. For example, tasks that involve the child creating a deception, in other words creating a false-belief, can be passed by children at around $2\frac{1}{2}$ years old, demonstrating that children younger than 4 years have some understanding of false-belief (Chandler, Fritz, & Hala, 1989). Concerns have also been raised about the test-retest reliability of ToM measures (Mayes, Klin, Tercyak, Cicchetti, & Cohen, 1996). Although reliability research for tests of ToM is somewhat limited, the handful of studies that have examined the psychometric properties of false-belief tests have reported mixed findings in relation to test-retest reliability; ranging from unacceptable to good reliability (discussed in more detail in Chapter 5 of this thesis) (Charman & Campbell, 1997; Hughes, et al., 2000; Hutchins, Prelock, & Chace, 2008; Mayes, et al., 1996). Concerns about reliability and validity of ToM measure have led to the development of ToM task batteries and scales (Muris, et al., 1999; Wellman & Liu, 2004). These more recent ToM tests were developed to measure a proposed developmental progression in understanding of mind and to provide a broader picture of children"s understanding of mind, rather than merely the ability to reason about false belief. Aside from concerns about the measurement of ToM, the nature of ToM is strongly debated. The most influential theories are presented here.

Theories of Theory of Mind

The focus of much ToM literatures has been the nature of ToM; including what it is, how it develops, and what brain region is engaged when a person uses his/her ToM . The consistent finding of a delay or absence of ToM in individuals with autism has inspired theories about autism and in turn theories about the nature of ToM (Lewis & Carpendale, 2002). Although theories about ToM differ greatly from one another, they can be generally reduced to being either theories that consider ToM to be a cognitive ability or theories that see ToM as an artefact of social interaction.

Within the cognitive theoretical framework there have been two dominant approaches to ToM and a third, less acclaimed approach. The first perspective that can be considered to be cognitive in nature, often referred to as the „Theory-Theory", views ToM as a learned ability. This account takes the view that ToM understanding is gradually acquired by children through the use of a trial and error learning method (Wellman, et al., 2001; Wellman & Liu, 2004). This perspective is Piagetian in nature, marking developmental milestones through the early years of development beginning with an understanding of desire (people want) around the age of 2 years, progressing to an understanding of belief (people think) around the age of 3 years, and eventually an understanding of false belief (people"s though can be inaccurate) around the age of 4 years. The second takes an innate perspective, which postulates that there is a mental module that mediates social functioning (Leslie, 1994). Central to this perspective is the Theory of Mind mechanism (ToMM). The argument for a module, or mechanism, stems from the seemingly automatic and fast processing of social stimuli during person-person interactions. Some have argued that this module is missing in Autism, thereby causing the social deficits of the disorder (Baron-Cohen, Leslie, & Frith, 1985). The third account has proposed that the type of reasoning involved in false-belief tasks is at the core of children"s pattern of emerging ToM (Peterson & Bowler, 2000; Riggs, Peterson, Robinson, & Mitchell, 1998). In particular, reasoning about counterfactual information is particularly difficult for young children. This theory has limitations for explaining patterns of ToM performance in children with autism; details of which will be discussed later in this chapter.

In contrast to the cognitive perspectives, the social interaction perspective is best thought of as an experiential account of mind. A number of theorists fall under this broad umbrella, some view ToM as an outcome of social interaction (Hobson, 2004), whereby children are thought to directly experience others‟ minds and in this way learn about people. While others take issue with the intellectualisation of social interaction, and argue that the concept of ToM is unnecessary (Costall & Leudar, 2004; Reddy & Morris, 2004; Shanker, 2004; Williams, 2004). Nevertheless, both approaches argue that a lack of engagement with others‟ minds is characteristic of children with autism. Hobson, (2004) argued that a lack of engagement with other‟s minds prevents learning in the social domain and subsequently delays or prevents the development of ToM (Hobson, 2004). In contrast, Williams (2004) argued that people with autism only have theories about other‟s minds, theories which are used to make sense of people‟s behaviour in hindsight, but are not useful in fast paced real world interactions.

Following is a detailed examination of these perspectives with respect to findings from the literature.

Cognitive Theories

'Theory Theory'

This cognitive theory of ToM assumes stages of development. These stages are described as analogous to a scientist developing a theory. The child‟s is thought to progress through a series of theories, each which advances on the previous theory, until eventually reaching a sound theory that explains the world most accurately (Astington, Harris, & Olson, 1988; Wellman, et al., 2001; Wellman & Liu, 2004). This theory has been termed the „Theory-Theory‟ and it draws its support from the apparent shift in ToM ability between the ages of 3 years and 4 years. Wellman‟s (1990) account suggested that around the age of 2 years children possess a desire-based psychology, that is, they understand people‟s actions in terms of what people want. The theory posits that through experiencing failures of a purely desire-based psychology, children evolve a more sophisticated theory, which encompasses the person‟s beliefs as well as desires, and this is what we see emerge around the age of 4 years. Recently, Wellman and Liu (2004) investigated these proposed stages using a ToM scale that included a single item to tap each developmental stage. The five-stage model begins with an understanding of desires, followed by an understanding that people can have differing beliefs, followed in turn by an understanding that seeing leads to knowing, then leading eventually to an understanding that people can have false-beliefs, and finally the understanding that people can feel one way but display a different emotion. A detailed account of this scale and Wellman and Liu‟s findings, are not presented here, because these are presented in Chapter 5 of this thesis. Wellman and Liu suggested that each earlier stage acts as a scaffold, preparing children for the next more complex stage in their developing understanding of mind.

Criticisms of the „Theory-Theory‟ have been made by both alternative cognitive accounts and contrasting social interactionist accounts of ToM. The primary criticism from both alternate cognitive accounts and social interactionist accounts has been that the „Theory-Theory‟ conception of ToM intellectualises a process that occurs largely outside cognitive awareness. The social interactionsist account has asserted that children and adults do not actively engage in the process of thinking about other‟s minds while interacting with them, instead people experience others by interacting with them (Costall & Leudar, 2004; Leudar & Costall, 2004; Williams, 2004). For example, a child does not think “my cuddle made Mummy happy” while interacting with his/her mother. Instead the child experiences his/her mother‟s reaction (feeling happy) to the embrace via the mothers expression of the emotion. Similarly the, ToMM account proposed that the processing of social information is faster and more automatic than is possible under the scope of the „Theory-Theory‟ (Friedman & Leslie, 2004; Leslie, 1994).

Theory of Mind Mechanism

The Mental Module account of ToM, put forward by Leslie (1994), positions children‟s understanding about the world in an innate mental module. Leslie (1994) described the characteristics of this module as follows:

1. Social interaction is time-pressured; hence a mechanism that processes social information must be able to do so rapidly.
2. A theory of mind mechanism (ToMM), "which is post-perceptual, operates spontaneously, and is domain specific, and is subject to dissociable damage – in limit, modular" (p. 214), is active during social interaction and specifically processes social cues, from a variety of sensory modalities.
3. Then the ToMM utilises a system that forms representations (proprietary representational system) to form representations of other‟s mental states.
4. "ToMM forms the specific innate basis for our capacity to acquire a theory of mind." (p. 214)
5. "ToMM is damaged in childhood autism resulting in its core symptoms and impairing these children‟s capacity to acquire a theory of mind." (p. 214)

In short, the ToMM is the means by which the brain recognises and then processes social information and that the proprietary representational system uses information from the ToMM to form meta-representations. As stated previously, meta-representation is thinking about thinking. Leslie (1994) argued that the earliest form of meta-representation is pretend play, whereby children demonstrate the capacity to both generate and understand pretence, or a sharing of thoughts about an object. Leslie proposed that any cognitive theory of ToM must be able to account for pretend play, which begins to emerge in children at around 18 months of age.

While the key elements of the theory, listed above, have often been supported by data from the literature, alternate accounts of ToM can also explain much of the data. An example of the way in which data from the literature can be used to support the ToMM view follows. After which, an alternate account is presented. Take for instance the fourth of Leslie"s propositions, that the ToMM is the means by which children acquire an understanding of minds. This proposition is supported by findings of very early attention to social stimuli by infants. It can certainly be argued that this very early preference for social stimuli reflects an innate drive or mechanism that seeks social information (i.e., aToMM). For example, very early in their lives babies tend to attend to human faces and voices and later as toddlers display joint-attention behaviours (Dawson, Toth, Abbott, Osterling, Munson, Estes, & Liaw, 2004). Joint-attention behaviours include pointing to share attention or interest in stimuli, following the gaze of another person to determine what they are attending to, and attending to the facial expressions of a caregiver to determine how to evaluate novel stimuli. These behaviours are readily observed in very young children, well before a basic ToM becomes evident. Flavell and Miller (1998) summarised this phenomenon eloquently "Human infants do indeed seem to be built with these two developmentally useful properties: They are impelled to attend to and interact with other people and they impel other people to attend to and interact with them." (p. 858). Furthermore, support for the modular view of ToM comes from findings of impoverished joint-attention behaviours in children with autism (Dawson, et al., 2004). Taken together these findings support the argument for an innate mental mechanism, which appears to be damaged in children with autism (as per point 5 of Leslie"s theory).

It is argued here, however, that there are problems associated with the use of some joint-attention measures in populations with autism. Primarily these are related to the salience of the stimuli used. That is, the degree to which the stimuli draw children‟s attention. Although there is certainly a difference in degree of social engagement between typically developing children and children with autism, an undetermined proportion of this difference might be attributable to differences in stimulus intensity. For example in Dawson et al.‟s study, joint-attention deficits were established through comparison between response to social and non-social stimuli. This comparison intended to show that children with autism demonstrate a specific deficit in attending to social stimuli. As expected, the findings indicated that, compared to developmentally delayed children and typically developing children, children with autism responded less to social stimuli. The finding, however, that all groups attended more often to the non-social stimuli suggests that stimuli salience may have played a role in the level of response among all children. Stimuli salience may have impacted on response to stimuli differently in the experimental groups. The non-social stimuli used by Dawson et. al. included a timer beeping, a phone ringing, a whistle blowing, and a tape recording of a car horn. These stimuli may be better at gaining children‟s attention than the social stimuli utilised in the study, which included humming a neutral tone, calling the child‟s name, snapping fingers, and patting hands on thighs. Salience might not only be associated with volume (loudness) of the stimuli but also with the acoustic properties of tones such as frequency or pitch. The reason for a difference in pattern of attention between typically developing children and children with autism is unclear. It might be that the social stimuli used fell below some threshold of awareness for the children with autism or that the social stimuli were perceptually more complex in nature (i.e., more difficult to process). Sensory processing differences present in children with autism might explain differences in early social processing and attention in these children (Bogdashina, 2005). To summarise, Leslie‟s proposition of a damaged or missing ToMM in autism is certainly supported by data of early differences in social attending between children with and without autism. The data can, however, also be used to support a theory of sensory processing disruption in autism.

Another example of the way in which data ambiguously add to the debate of theories of ToM is presented in relation to Leslie"s fifth proposition of the ToMM theory. According to this proposition people with autism should not develop a ToM. There is however, evidence of delayed instead of missing ToM in children with autism (Happé, 1995). Nevertheless, evidence that ToM can develop in people with autism is not entirely problematic for the ToMM account of autism. The response of the proponents of the mental module position, to evidence of ToM as it is inferred from false-belief task success, has been that when individuals with autism give correct answers to false-belief questions and other tests of ToM, the process culminating in that answer is not like that of people without autism. Leslie argued that rather than a fast, automatic, effortless response, the responses of people with autism are „hacked out" in a far more calculated way. The term „hacked out" is used as a way of emphasising the amount of effort required to think about the mental worlds of others. This, then, is how people with autism learn to compensate for their ToMM deficit.

In order to substantiate this claim, research needs to find that the responses of people with autism are slow for ToM questions specifically. It might be the case that people with autism generally respond more slowly to questions regardless of whether or not they involve mental states. For example, questions that do not require reference to other minds but that can be answered in a fast, automatic, and effortless way. Although limiting the real-life relevance, this type of research needs to rely on laboratory-type tasks, mainly because developing real-time interaction tasks for this type of research would need to overcome difficulties of equating stimulus complexity. That is, tasks that use non-social stimuli need to be matched for complexity to tasks that use social stimuli. While laboratory tasks, such as the false-belief task, involve relatively simple stories that can be matched for language complexity and cognitive demand, real social interactions involve processing auditory, visual, and semantic information simultaneously and cannot be easily replicated with non-social interactions; even less so when adding the extra requirement of an automated modular process. Limitations of laboratory studies to answer this research question include the relevance of findings made under such constrained conditions for real world interactions. Findings from laboratory studies that use impoverished social stimuli may not extend to real social interactions. Without this form of timed response research it is difficult to draw conclusion about the way in which people with autism process social information.

Despite a lack of research, limited support for the proposition of a laborious process, although not intentional, comes from bibliographies of high functioning adults with autism (Williams, 2004). As summarised by Williams, these accounts all contain reference to learned rules of social engagement that do not often help with real-time social engagement. The authors of these accounts have explained how they reason in social situations based on the rules they have learned, but that this process is often too slow to aid effective social interaction and that most of the understanding occurs in hindsight. Williams has, however, been opposed to the cognitive accounts of ToM, and would likely suggest that, although the biographical accounts support the idea that ToM in autism is nothing like ToM in typically developing individuals, this does not indicate that ToM is an automatic cognitive process. As will be discussed later, Williams proposed instead that ToM is a social experiential phenomenon.

As illustrated by the above examples, data can often be used to support the propositions of the ToMM as well as that of contrasting theories. Despite this, the primary limitation of the „mental module" account as an explanation for autism is its inability to explain non-social symptoms of autism. If a ToMM module was only thought to underpin social development, an absence of such a module could conceivably explain the impaired language and social functioning characteristic of autism. Another facet or explanation remains necessary for those symptoms of autism left unexplained; specifically the core symptoms of repetitive behaviours and inflexible routines. While the aforementioned examples of joint-attention and fast automated processes can be used equally to support the ToMM account and alternative accounts of ToM from a social interactionist perspective, the social interactionist accounts are better able to explain the remaining core symptoms of autism. This is because, in order to explain all the core symptoms of autism the ToMM account needs to invoke a means by which a damaged ToMM results in repetitive behaviours and inflexible routines. Any theory that can effectively account for all the symptoms of autism more minimally (sensibly, falsifiably, and accurately) ought to be considered a better theory than a counter theory that proposes a myriad of disturbances to explain the same set of symptoms.

Counterfactual Reasoning

The third cognitive theory posits that preschool aged children have difficulty with a specific type of reasoning, namely reasoning about counterfactual situations. Counterfactual reasoning involves reasoning about events or situations that contrast with present reality. Key researchers in this area have argued that this type of reasoning forms the basis of false-belief reasoning (Peterson & Bowler, 2000; Riggs, et al., 1998). These theorists proposed that patterns of findings about early ToM in relation to success or failure on false-belief tests can be explained from a counterfactual reasoning position without needing to impute a specific difficulty with understanding other people‟s minds.

Data from this line of research have been mixed, as described below. Some findings have supported the notion that children with and without autism have difficulty generally reasoning about counterfactual events rather than a specific deficit in an understanding of mind, while other data have been ambiguous. Early research examining the subtractive reasoning component of counterfactual reasoning found a general difficulty with counterfactual reasoning in typically developing children (Riggs, et al., 1998), while later research found a that only children with autism were specifically impaired in reasoning about false-beliefs (Peterson & Bowler, 2000).

The authors of these studies proposed that ToM tasks involve a particular type of hypothetical reasoning which they say is "from a logical point of view…counterfactual in the sense that a known fact is contradicted" and "from a cognitive point of view…subtractive in the sense that the known fact has to be ignored or overridden as if it were not represented." (Peterson & Bowler, 2000, p. 392). To test this claim both studies presented children with two stories followed by a false belief question for one story and a subtractive reasoning question for the other story. Although the studies did not utilise the typical Sally-Anne story, they used stories constructed in the same way. The subtractive reasoning questions asked what something would be like if an event had not taken place; for example "Where would the ball be had Sally not moved it?"; where the corresponding false-belief question might be "Where will Anne look for the ball?" Peterson and Bowler asserted that the subtractive reasoning question is the supposition that a child needs to generate in order to answer the false-belief question correctly. While subtractive reasoning success and false-belief success were related in both typically developing children (Peterson & Bowler, 2000; Riggs, et al., 1998) and children with autism (Peterson & Bowler, 2000), children with autism found the false-belief tasks particularly difficult (Peterson & Bowler, 2000). Peterson and Bowler proposed this might be due to a difficulty with generativity (generating a spontaneous response). That is, that in order to answer the false-belief question children must generate the „if not...then...‟ supposition spontaneously, whereas in answering the subtractive reasoning question this supposition is already given. The authors argued that a lack of generativity is a general problem for children with autism and cite examples in relation to other areas such as play and language, where children with autism can often complete tasks when prompted but have difficulty doing so spontaneously. The authors recognised that their findings could also be accounted for by Leslie and Thaiss‟s (1992) ToMM account of ToM.

Other research findings have contradicted the conclusion that young children have difficulty reasoning about counterfactual situations (Harris, German, & Mills, 1996). Recently

German and Nichols (2003) attempted to make sense of these contradictory findings. They questioned whether the difficulty with counterfactual reasoning demonstrated by Peterson, Riggs and colleagues (Peterson & Bowler, 2000; Riggs, et al., 1998) was associated with the counterfactual nature of reasoning tasks or whether it was associated with task complexity. To test this, German and Nichols created three counterfactual reasoning tasks of varying complexity. Complexity for these tasks was related to the number of causal links children had to reason about backward, in order to answer the test question. Children were all presented with the same stories consisting of a chain of linked events (e.g. woman planted flower, woman was happy, woman called her husband, husband opened door, dog escaped, dog ran around garden, dog trampled newly planted flower, woman was sad). Reasoning complexity was varied by asking children how the woman would feel if: 1) she had not called her husband; 2) if the dog had not escaped; 3) if the dog had not trampled the flower. German and Nichols termed these long, medium and short chain inferences, respectively. In order to answer the test question correctly, children who were asked the long chain test question needed to infer: Had the woman not called her husband, he would not have opened the door, the dog would not have escaped, and would not have run around, it would have not trampled the flower, therefore the woman would still be happy. The medium chain inferences were: had the dog not escaped, it would not have run around the garden, and would not have trampled the flower, therefore the woman would still be happy. The short chain inferences were: had the dog not trampled the flower, the woman would still be happy. Results from the study indicated that short chain inferences were significantly easier than medium and long chain inferences among 3 and 4-year old children. No difference in difficulty was found between medium and long chain inferences. Furthermore, false-belief task success was positively related to success with medium and long chain inferences but not related to short chain success.

In contrast to Peterson and Bowler‟s conclusion that children‟s false-belief difficulty is a result of a general difficulty with counterfactual reasoning, German and Nichols concluded that preschool aged children have no specific difficulty with counterfactual reasoning. Instead, they proposed that this ability is masked by task artefacts, namely inference complexity. As argued by German and Nichols, it is likely that similar cognitive demands associated with both counterfactual reasoning tasks and false-belief tasks might account for the observed relationship between these measures. Problematic for this theory is that short chain inference tasks, which are most closely associated with false-belief reasoning because each tasks requires a single causal inference, were not found to be related to false-belief tasks. German and Nichols proposed that this might be due to added cognitive complexity inherent in answering false-belief questions. They proposed two possible accounts of processing demand. The first is in relation to inhibiting a response about the current state of the environment in place of a response reflecting the past state of the environment. While inhibitory processes have been implicated in false-belief reasoning, and are probably important for correct responding (Joseph & Tager-Flusberg, 2004; Müller, Zelazo, & Imrisek, 2005), such processes would also be necessary for the short chain tasks. To explain why inhibition of a „current state‟ response was not as important for short chain responding, German and Nichols posited that with each step of inference an extra step of inhibitory processing is required and that with each inference this becomes increasingly difficult for young children. A second, more plausible, explanation implicated working memory load. It is also likely, as mentioned by the authors, that both inhibitory processes and memory load contribute to task performance. Unfortunately, German and Nichols did not measure children‟s cognitive, verbal or executive abilities, measures which would help contribute to the understanding of the relationship between counterfactual reasoning and false-belief tasks.

In addition to the cognitive theories that have dominated ToM research, social developmental theories have also been put forth. This chapter will now examine an alternative account of ToM from a social interactional perspective, in order to demonstrate an alternative explanation for both typically developing ToM and the relationship between the core symptoms of autism and a delayed and different ToM capacity.

Social Theories

Research has found that often ToM ability in children does not necessarily make them socially more able (Reddy & Morris, 2004). This is a puzzling finding considering that ToM ability is treated as marking a developmental stage deemed important by most ToM researchers. Moreover, the disconnect between social ability and ToM is especially perplexing since the lack of ToM ability in autism has been widely regarded as the underlying cause of their social deficits (Leslie, 1994; Baron-Cohen, Leslie, & Frith, 1985). How can a fundamental skill be responsible for social aptitude in one population but not in another? While theoretically interesting and enticing, mainstream ToM explanations have, more or less, to the exclusion of all other explanations treated ToM as a cognitive phenomenon or skill set. While these theories generally consider social input as the means by which ToM matures, cognitive theories treat social stimuli as a catalyst in their proposed processes of developing an understanding of mind. Cognitive approaches have to some extent disregarded social environmental influences on the development of ToM, as well as ignored the real world implications of having a ToM, and thus resulted in a limited understanding of social development and indeed a limited understanding of the development of ToM. An alternative explanation of ToM deficits comes from a social interaction perspective and this will be discussed now.

Triadic Interaction

Social interactionist approaches have focused on the typical types of interactions children have in the early years of life. Carpendale & Lewis (2004) stated that “triadic interaction between the child, another person, and the world is essential in the development of knowledge”

(p. 85). Triadic interactions involve joint attention behaviours, which Hobson (2004) declared is how children learn that people have thoughts and beliefs that differ from their own thoughts and beliefs. Joint attention is best thought of as a set of behaviours that function to align the attention of two or more people to some aspect of interest; for example orienting to the focus of another person"s attention by following that person"s gaze or point or initiating shared attention by pointing. Hobson suggested that an inability to relate affectively with others leads to a lack of joint attention in early childhood, which in turn results in a lack of learning about the world and the social world in particular, as seen in childhood autism. Hobson presented evidence that social learning deficits similar to those found in autism are found among non-autistic children born blind and children who are deprived of early social interaction in severe neglect situations. These findings support Hobson"s theory in so far as they show that sensory deprivation leading to a lack of ability to engage in triadic interactions can lead to social deficits similar to those found in autism.

Further support for the theory that impoverished joint-attention experiences results later in the core social and language deficits of autism comes from recent research by Delinicolas and Young (2007) who demonstrated that joint-attention behaviours in children with autism *were* most strongly associated with core deficits that were theoretically linked to joint-attention, such as language and social relating. Delinicolas and Young argued that while other research has demonstrated poor joint-attention in autism, the apparent relationship may simply be a function of co-occurrence of these symptoms. That is to say, a lack of joint-attention has not been demonstrated to have a causal relationship with diminished social and language abilities and instead all may simply be products of the underlying cause of autism. By comparing the strength of the correlations between joint-attention and the core symptoms of autism, Delinicolas and Young were able to demonstrate that joint-attention was indeed related to the theoretically linked core symptoms of language and social relating more so than the core symptoms of autism not theoretically linked, such as stereotypical behaviour.

Nonetheless, the cause of impoverished triadic interactions and joint attention in autism is unclear. There are two plausible and competing accounts. The first, as described earlier, is a missing ToMM that when functioning orients children to social stimuli. The second is that sensory processing difficulties make orienting to and processing social stimuli difficult for children with autism (Bogdashina, 2005). The later account is supported by findings that children born deaf to hearing parents are able eventually to learn to compensate for their disability and do develop a delayed ToM (Peterson & Siegal, 1995; Peterson, Wellman, & Liu, 2005). Children with autism, on the other hand, are arguably always hindered in their learning and in interacting with others. The ongoing nature of social difficulties in autism can be accounted for by an underlying sensory processing disorder, a disorder that initially caused the developmental disruption and continues to disrupt interaction.

Triadic interaction theory is intuitively appealing because it does not rely on specialised innate mechanisms nor does it complicate social interactions by inferring hidden intellectual processes. Furthermore, while triadic interaction theory does not purport to explain all the symptoms of autism, taken together with a view of autism as a sensory processing disruption or difference, it is able to account for all the core symptoms of autism. That is because, theoretically, a sensory processing disruption could also explain repetitive behaviours and inflexible routines. Brain development involves, and therefore learning about the environment involves, the formation of synaptic pathways (Medrihan, et al., in press). Medrihan et al., found a specific protein deficiency in mice that disrupted normal synaptic functioning. They stated that a similar brain-based protein deficiency in autism is likely to be responsible for the synaptic dysfunction found in autism. Where the typically developing brain automatically filters and stores sensory information received from the environment, this process is believed to be disrupted in people with autism; a disruption which, according to Medrihan et al., is due to an excitatory-inhibitory imbalance in synaptic processes. A disruption of this kind can explain all of the core symptoms of autism. For example, if this orderly process is disrupted it might be that the result is an inability to attend to relevant stimuli in the environment. Theoretically this could result in the brain being overwhelmed by irrelevant stimuli; in turn, manifesting in an individual attending exclusively to a particular stimulus in order to relieve or avoid overstimulation. In autism this might look like repetitive and stereotypical behaviours. Similarly inflexible routines might be necessary for an individual whose brain does not file new information efficiently or effectively. It is conceivable that any departure from normality could be disturbing for a person who is not able to make sense of new information.

This current thesis does not, however, concern itself with sensory or neural processes in autism. The theoretical process by which a sensory or synaptic processing disorder might explain autism is presented only by way of demonstrating alternative possibilities to the commonly held views of ToM. Chapter 2 of this thesis will examine more closely the correlations between ToM, language, reasoning, cognitive, and social development.

CHAPTER 2 – Theory of Mind and its Correlates: The Relationship between ToM and Cognitive, Language and Social Development

This chapter will explore the relationship between ToM and aspects of child development that are theoretically linked to it. Published research in this field has generally been linked in one way or another to key theories or theorists of ToM. While not all research has been motivated in this way, researchers in this domain have typically subscribed to one particular theoretical position. Subsequently, findings from the research have been used to differentiate between aspects of competing theories. Data have often supported a particular theory, but not exclusively so. Although findings have often been robust, ambiguity has also existed in relation to some aspects of the data. As can be reasonably expected, researchers have generally provided explanations for ambiguous data that fit with the theory to which they subscribe; or alternatively researchers have proposed uncontrolled factors that might explain unusual data. Nonetheless, alternative explanations for confusing data have sometimes been equally plausible. In short, this chapter will present key findings from individual differences research, in the ToM domain, and where data fit alternative explanations, these will be discussed.

A number of theories about the nature of ToM developed from the early methodology of the ToM research domain; which centred on determining group norms (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). This normative group data raised questions about why ToM does not develop in some populations and what children"s conceptual understanding is like before ToM develops. To better understand these group differences, individual differences research was conducted to examine other areas of development in relation to ToM ability. The relationship between ToM and other aspects of child development has been thought of in a number of ways. Factors have been discussed as: contributing to the development of ToM; contributing to the development of ToM while also contributing to performance on tests of ToM; related to performance on ToM tests, but not contributing to the development of ToM; and developing alongside ToM but in no way related to ToM. Moreover, factors have been researched within a number of developmental domains. This chapter will organise research findings by developmental domains. Specifically, the chapter will begin with an examination of the relationship between ToM and cognitive development. The chapter will then address the complex relationship between ToM and language; including the role of linguistic ability and early communicative experiences. The chapter will conclude with the aims of this thesis.

Theory of Mind and Cognitive Ability

ToM has been studied in relation to cognitive abilities in typically developing children and children with autism. Cognitive factors have generally been discussed as contributing to performance on false-belief tasks. Primarily, this stems from the fact that ToM research has relied heavily on empirical research designs; which for the large part have used false-belief type tasks to assess children‟s understanding of the mental lives of others (Wellman, Cross, & Watson, 2001). Tasks believed to measure ToM are inherently cognitive in nature; as opposed to observational. That is to say, tasks test children. It is not possible to directly measure people‟s cognitive processes. Instead researchers devise tasks that best isolate a particular skill or skill set according to accepted criteria for validity. While indirect measurement may not be the most precise form of measurement, it ought not to be problematic if the skills required to complete the task are universal or rudimentary. For instance a task that uses printed text to measure language comprehension relies on a properly functioning visual system; although the visual system itself is not likely to be responsible for text comprehension. While a damaged visual system might impede task performance, it would not be implicated in language comprehension. Similarly, ToM tasks typically rely on rudimentary and universal skills such as memory, language comprehension, and so forth. However, while the involvement of unrelated simple cognitive abilities are ordinarily unproblematic in research with adults or older children, research with younger children and clinical populations cannot ignore the impact of cognitive abilities because these are still developing at differential rates both within and between individuals and in some cases development is impaired.

In the broad field of ToM research a number of cognitive factors have been suggested as either linked to ToM or responsible for some of the symptoms of autism that cannot be explained by a ToM account alone. Frith (Frith, 1989; Frith & Happé, 1994) put forward central coherence as an account for aspects of autism that are not otherwise explained by a ToM account. By this account, processing stimuli as individual elements rather than as a coherent whole is used to explain non-ToM related characteristics of autism. However, although some researchers have found such deficits, independently of ToM, (Pellicano, Maybery, Durkin, & Maley, 2006), others have reported a direct link between ToM and central coherence (Jarrold, Butler, Cottington, & Jimenez, 2000). ToM has also been studied in relation to aspects of memory, for example working memory in autism (Ozonoff & Strayer, 2001) and episodic memory in relation to performance on ToM tests (Perner, Kloo, & Gornik, 2007). Although the aforementioned areas of research have been guided by sound theoretical associations, the most commonly implicated factors in ToM have been in three areas pertinent to both development in autism and typical development; namely domain general reasoning, executive functions, and language. These findings will now be reviewed.

ToM and Reasoning

ToM tasks might be thought of as reasoning tasks that involve reference to mental states. Researchers have, therefore, examined how reasoning about mental states is related to equivalent reasoning about non-mental phenomenon (Amsel, Trionfi, & Campbell, 2005; Grant, Riggs, & Boucher, 2004; Leslie & Thaiss, 1992; Peterson & Bowler, 2000; Riggs, Peterson, Robinson, &

Mitchell, 1998; Zaitchik, 1990). This body of research has produced a range of interesting data, which will now be discussed.

The earliest research found that mental state reasoning was no more difficult than reasoning about other representational phenomenon (false photographs); at least not for typically developing children aged between 3-years and 5-years (Zaitchik, 1990). The false-photograph tests, used by Zaitchik, were similar in structure to standard false-belief tests. For both types of tests an initial state of affairs is changed and children are asked to reason about a representation of the original state. The representation is either a mental state (i.e., a belief) for false-belief tasks or a photograph for false-photograph tasks. By correctly responding to the test questions, children either demonstrate an understanding of the mental representation or the pictorial representation, both of which no longer represent reality. Therefore, Zaitchik concluded that young children‟s difficulty with ToM tasks was the result of a general deficit in understanding representation.

Leslie and Thaiss (1992) replicated Zaitchik‟s findings with typically developing children, but not with children with autism. Leslie and Thaiss found that children with autism performed significantly better on false-photograph tasks than on false-belief tasks of similar difficulty. From this finding, Leslie and Thaiss concluded that different cognitive mechanisms underpin reasoning about mental states and reasoning about other non-mental representations. Leslie and Thaiss argued that their findings supported Leslie‟s account of a ToM mechanism (ToMM) which is damaged in people with autism. However, an alternative explanation might be that the photographic representation is more accessible than the mental representation (belief) for children with autism. The photograph of the previous state of affairs is on the table in front of the child, albeit face-down, whereas mental representations, in this instance another person‟s beliefs, may not be quite as salient, nor tangible.

While an understanding of the nature of representations might underpin successful reasoning about both false-beliefs and false-photographs, both these forms of reasoning also fit into a broader form of reasoning about counterfactual information. Peterson and Bowler (2000) reported findings that suggest that the type of reasoning involved in ToM tasks is comparable to the type of reasoning involved in other tasks; tasks that do not involve mental states, but are similarly counterfactual in nature. These authors examined the abilities of young children, children with autism, and children with severe learning difficulties by presenting children with two standard false-belief stories and asking children either a false-belief question (FB) or a subtractive reasoning (SR) question. The SR question differed from the FB question in that instead of asking "Where will Anne look for the ball?" the question was "Where would the ball be had Sally not moved it?" In total, children were asked one SR question and one FB question. Peterson and Bowler did not use the standard Sally-Anne stories; instead they used comparable stories, for which the number of character actions and story complexity were equivalent to a standard Sally-Anne task. Peterson and Bowler proposed that the SR question was the supposition that the child would otherwise need to generate in order to answer the FB question correctly. In other words, to answer the false-belief question correctly, the child must first consider that the event did not take place for the other person, who was absent and did not witness the change of state.

In line with Leslie and Thaiss" findings, Peterson and Bowler also found that children with autism have specific difficulty with reasoning about mental states. Children with autism were more likely to answer the SR questions correctly than the FB questions. On the other hand, the majority of typically developing children either correctly answered both SR and FB questions or were unable to answer either question correctly. Similarly, children with severe learning difficulties found both tasks equally difficult. Once again, the underlying cause of the specific difficulty with reasoning about mental states shown by children with autism remains unclear.

While Leslie‟s (1994) ToMM is able to explain these differences in light of a damaged mental module, others have logically proposed that a specific difficulty in understanding mental states might stem from differences in early social experiences inherent in this population (Peterson, 2002).

Additionally, Peterson and Bowler‟s study examined the role of verbal ability in successful FB and SR reasoning. The authors presented descriptive statistics in relation to chronological age and verbal mental age for children who passed and children who failed the FB and SR questions. Peterson and Bowler found that in general children who correctly answered FB and SR questions had a higher chronological age and higher verbal mental age. Unfortunately, however, the authors did not compare the mean verbal ability of children across tasks. For example they did not compare whether there was a difference in verbal ability required to pass FB questions, in comparison to SR questions. It is, however, still possible to examine the pattern of results; in lieu of inferential statistics. A clear pattern emerges from these data. For typically developing children, similar levels of verbal ability (mean verbal mental age equivalent to 5-years old) were necessary for success on both FB and SR questions. On the other hand, it appeared as though children with autism required a higher verbal ability to pass FB tests (mean verbal mental age equivalent to 7.6 years old) than that required to pass SR tests (mean verbal mental age equivalent to 6 years old). Children with autism who did not pass the SR tasks had, on average, a verbal ability equivalent to 3.8 years and those who did not pass the FB tasks had a verbal ability, on average, equivalent to 4 years of age. Although not equivalent, this pattern of results is somewhat consistent with a large meta-analysis that examined the role of verbal ability in relation to ToM for children with autism and found that children with autism generally do not succeed on ToM tests until they have a verbal ability approximately equivalent to 9-years-old (Happé, 1995).

Peterson and Bowler proposed that the pattern of performance of children with autism reflected a more global difficulty with “generativity”. The term “generativity” refers to the ability to spontaneously generate a required behaviour. The authors proposed that the difficulty evidenced by children with autism on the FB question may be attributable to the fact that, when answering the FB question children had to generate a supposition spontaneously; a supposition which was provided in the form of the test question in the SR task. Peterson and Bowler referred to evidence of a wide spread difficulty with generativity in autism; evidenced by findings in other areas such as play and language, where children with autism can often complete tasks when prompted but have difficulty doing so spontaneously.

To summarise, tasks that assess domain-general reasoning and those that assess reasoning specifically about mental states appear to place similar demands on typically developing children, while presenting children with autism with unique challenges. A specific difficulty with reasoning about mental states may well be the result of a damaged mental module in autism or impoverished learning about social phenomena. Whatever the cause of the increased difficulty of the mental state tasks for children with autism, the relationship between domain-general reasoning ability and reasoning about mental states is likely to be the result of more general executive functions that mediate both types of reasoning. The role of executive functions in ToM reasoning will now be discussed.

Executive Functioning

Correlations between Executive Function and ToM

Executive functions are described as a range of higher order cognitive skills, including inhibition of dominant but not necessarily correct response, impulse control, planning, organised search, and flexibility of thought and action (Ozonoff, Pennington, & Rogers, 1991). Executive functions are impaired in a number of clinical populations, including autism (Pennington &

Ozonoff, 1996). While research has separately examined executive functioning and ToM in children with autism, Ozonoff et al. (1991) proposed these abilities might be related in this clinical cohort. Specifically, Ozonoff and colleagues proposed that both ToM deficits and executive function deficits were symptoms of underlying neurological damage in autism and would, therefore, be found to be correlated. In line with their predictions, Ozonoff et al. found that executive functioning and ToM were only correlated in the group with autism and not in a matched group of controls without autism. Subsequently, the authors concluded that in autism an underlying brain disorder affects both executive function and ToM.

Moreover, and in light of dominant research findings of a ToM deficit in autism, Ozonoff et al. also expected to find a highly prevalent ToM deficit and a less prevalent executive function deficit; instead they found the exact opposite. Using a composite measure derived from the raw scores of two common executive function tests (Tower of Hanoi and Wisconsin Card Sorting Test (WCST)) they found that 96% of participants with autism exhibited poorer executive functioning than the mean level of executive functioning of the matched control group. In comparison, on the ToM composite measure only 52% of the participants with autism scored more poorly than the mean score of the control group. Ozonoff et al. therefore proposed that a prefrontal impairment, common to clinical populations who evidence executive functioning deficits, could be responsible for some of the symptoms of autism.

An alternative explanation for the findings might be that the composite ToM measure was not a valid measure of ToM. The composite test score included raw scores from 14 measures of ToM; a standard false-belief test and other tests believed to measure understanding of mind devised by Baron-Cohen and colleagues (1986, 1989a, 1989b) in earlier research. Inspection of performance across these tasks reveals that participants with autism were not significantly impaired on six out of 14 of these tasks, but were significantly impaired on the remaining eight tasks. It is possible that these tasks had questionable construct validity; construct validity of ToM tasks will be examined in a later chapter. Although, Ozonoff et al. did not propose that construct validity of the measures was responsible for the variable ToM scores of participants with autism. The authors did, however, express concern over the lack of replication of Baron-Cohen"s earlier findings in relation to these measures.

Moreover, although Ozonoff et al. proposed that the relationship between ToM and executive functions was unique to autism, research findings have extended the relationship beyond autism. Research with typically developing children (Hughes & Ensor, 2005; Moses & Sabbagh, 2007; Müller, Zelazo, & Imrisek, 2005) and children with attention and behavioural problems but without autism (Fahie & Symons, 2003) has found correlations between executive function and ToM beyond shared maturational effects. Additionally, neuropsychological research using functional Magnetic Resonance imaging (fMRI) has found that executive functions contribute to the cognitive process underlying ToM reasoning in non-clinical adults (Saxe, Schulz, & Jiang, 2006). In light of mixed and sometimes contradictory findings, it is not surprising that researchers have raised concerns about discrepancies in the varied findings of executive function research in autism (Liss, et al., 2001). Therefore, a closer look at executive functioning in autism follows.

A Closer Look at Executive Functioning in Autism

Liss et al. summarised relevant literature and concluded that overall research findings indicated that people with autism do not have a global executive functioning impairment, but rather a specific tendency to perseverate and perhaps a planning deficit. To perseverate is to continue giving a certain type of response even when feedback indicates this response to be incorrect. Furthermore, Liss et al. claimed that executive function deficits are not as universal in autism as first thought.

In order to address these discrepancies, Liss and colleagues examined various aspects of executive functioning in high functioning children with autism and a clinical population without autism but with a developmental language disorder; that is children without a psychiatric diagnosis but with a language impairment. The groups were matched for mean full scale IQ and mean nonverbal IQ. All children were administered four measures of executive functioning, each measuring different components of executive functioning. Liss et al. listed these components as "planning, set shifting, perseveration, sustained and focused attention, and rapid retrieval of verbal information." (p. 263).

Liss et al. did not find executive function deficits in children with autism when compared to children with a developmental language disorder. The only significant difference between the groups was an increased tendency to perseverate in the group with autism, although this difference disappeared when the authors controlled for verbal IQ. In both groups full scale IQ was positively correlated to better performance on all executive functioning measures. In the group with a developmental language disorder all but one executive functioning measure, number of correct responses on the Wisconsin Card Sorting Test (WCST), were correlated with verbal and non-verbal IQ. For the group with autism, executive functioning measures were less correlated with verbal IQ measures, but still correlated with non-verbal IQ measures. In contrast to the claims of Ozonoff et al., Liss et al.‟s findings demonstrated that an executive functioning deficit is not unique to autism.

Liss et al. explained the differences in these findings in relation to verbal ability. The authors logically postulated that a tendency to perseverate may actually be an artefact of poor verbal comprehension. The WCST uses verbal feedback to indicate that a sorting strategy is incorrect. If the person completing the task does not attend to that feedback then they might continue with a previously used sorting strategy. To support this alternative explanation for perseveration in autism, Liss et al. cited Ozonoff‟s (1995) findings of significantly better performance on a computerised version of the WCST among people with autism. Some support for Liss et al.‟s proposition has come from Joseph and Tager-Flusberg‟s (2004) finding that, once verbal ability was controlled for, ToM no longer correlated with a number of executive function measures.

Joseph and Tager-Flusberg found predictable correlations between ToM, executive function, and symptom severity in autism, only before controlling for verbal ability. For the most part, executive function tasks were correlated with each other, with ToM and also to some extent with symptom severity in autism. Furthermore, ToM was correlated with all measures of symptom severity. Exceptions were found for only two measures, the Day-Night executive function task and the Social interaction symptom severity scale, which were not broadly correlated with other measures. A likely reason for this, as highlighted by Joseph and Tager-Flusberg, is that the distribution of the Day-Night task was positively skewed and was, therefore, less sensitive to individual variation.

Despite broad correlations between measures, once verbal ability was taken into account a limited number of correlations remained significant. Only two correlations remained between the four executive function tasks. The first was between a task measuring inhibition and working memory (the Knock-Tap) and a task assessing planning (Tower task) (.44) and the second was between a task measuring working memory (Total Span) and a task measuring inhibition and working memory (Day-Night task) (.39). Additionally, ToM remained correlated with the Knock-Tap executive function task (.48) and the communication component of the symptom severity measure (.64), as measured by the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). After controlling for verbal ability, significant correlations (range = .54 -.67) were also found among symptom severity measures across the three separate diagnostic domains (communication, social interaction, repetitive behaviours).

These findings highlight a number of issues relating to both the nature of the relationship between ToM and executive functioning, and also the measurement of the abilities.

Firstly, from the data it appears that the relationship between performance on ToM tasks and executive functioning in autism is not altogether the result of verbal ability mediating performance on both types of measures, as suggested by Liss et al. (2001). Liss and colleagues proposed that verbal ability likely influences performance because both ToM and executive functioning tasks require that children attend to and respond to verbal cues. A number of tasks used by Joseph and Tager-Flusberg had very limited verbal cues. Moreover, children were given training in the tasks before testing began in order to ensure that all children understood task instructions. Importantly, two of the executive functioning tasks used by Joseph and Tager-Flusberg did not rely on verbal cues during task completion (Knock-Tap and Tower). From this it appears that, although verbal ability mediates performance on both ToM and executive functioning tasks, an additional cognitive skill or skills, beyond language, may be important for both tasks. Perhaps of concern is that performance on Knock-Tap was the only executive function measure that correlated with ToM score once verbal age was taken into account. Poor reliability of other executive function measures might have obscured the relationship between those executive function tasks and ToM tasks.

Joseph and Tager-Flusberg‟s sketchy findings highlight the importance of using reliable tasks. Mixed findings are more likely to be a consequence of unreliable tasks rather than an unstable relationship between executive function and ToM variables. For example, one would expect a strong association between tasks utilising the same procedure but different stimuli. While this was the case for Word and Block Span tasks, it was not so for Knock-Tap and Day-Night tasks. The latter tasks require that children inhibit a salient response in light of the rule of the game (say/do the opposite). The main difference between these tasks is that one requires a verbal response (Day-Night) whereas the other requires a physical response (Knock-Tap).

Certainly, we would expect that the correlation between these structurally related tasks would be higher than between either of these tasks and any other measure. As previously mentioned, near ceiling performance on the Day-Night task was problematic for analyses involving this measure. This is unfortunate because interesting and relevant comparisons could have been drawn between the relationship between equivalent verbal and non-verbal executive tasks in relation to ToM. For example, comparing the correlation between ToM and Day-Night tasks with the correlation between ToM and Knock-Tap, once verbal ability had been accounted for, could have clarified the role of verbal ability and of executive functioning specifically working memory and inhibition.

Research exploring ToM and executive functions has used a wide range of tasks to measure these concepts. Commonly used executive function tasks and ToM tasks differ in degree of verbal instruction, complexity and structure, differences that exist both within and between task types. Task differences might contribute to the relationships found between target variables within a study but also across studies employing different measures. It is, in part, because of these differences that the specific roles of executive functions in performance on ToM tasks remain unclear.

It might well be the case that children perform more poorly on ToM tasks if they are not able to maintain attention or are unable to switch focus between contrasting pieces of information; skills considered to be executive functions. While attention to the story is generally assessed in standard false-belief tasks by way of memory control questions, a correct response to control questions does not eliminate the possibility that a child‟s attention might have waned during the task. Nevertheless, inclusion of these questions does to a large extent measure story comprehension and recollection. Potentially more important is the ability to switch attention to relevant information. For instance, to answer a false-belief question correctly, children need to stop thinking about where the item of interest actually is, in order to indicate the item‟s previous location. In order to better measure the role of executive functions in answering ToM questions it is necessary to devise tasks that can isolate understanding of mind and executive functions. This line of literature will now be reviewed.

Modifying Common Tasks to Isolate the Role of Executive Functions

A deficit in understanding others" minds, a specific deficit in executive functioning, or a combination of both might impact upon performance on ToM tasks. Up until now, the research discussed in this chapter has demonstrated an ambiguous relationship between ToM and executive functions. If executive functions do mediate ToM, this might help to explain findings that children younger than 4-years-old can pass certain tests of ToM (Chandler, Fritz, & Hala, 1989; Sullivan & Winner, 1993). To gain a clearer understanding of how young children and children with autism succeed on ToM tests or why these same children do not succeed on standard false-belief tests, studies have manipulated the level of complexity of false-belief tests and other deception-based ToM tests.

In one such study, Russell, Hala, and Hill (2003) modified a deception task, the windows task, to isolate the element of deception. The study aimed to examine whether executive function mediated performance on tasks that require that children create deception. Ordinarily the windows task employs a box with three solid sides and one transparent side. The box contains two compartments, which can both be viewed through the transparent side of the box. Only one compartment contains a treat. The child participant sits so that s/he can see into both of these compartments, while a competitor, who cannot see into the box, sits on the opposite side. To test the child"s ability to deceive, the child is asked to indicate the location of the treat to the competitor. If the child indicates the location of the treat, the competitor receives the treat. On the other hand, if the child tells the competitor to look in the empty compartment the child receives the treat. In this way the task requires the child to create a false-belief in another person.

Russell et al. modified the box by introducing an automated dispensing process. Consequently, the child only had to press a button to respond, rather than give a verbal response. Russell et al. reported two studies, the first compared typically developing children aged three and four years, the second compared children with autism and children with moderate learning difficulties. Children in each group were randomly divided into three conditions: No-Opponent condition, Opponent-non deceptive condition, and an Opponent-deceptive condition. In the No-Opponent condition, the child merely pressed the button corresponding to the empty side to receive the treat from the other side. In the Opponent-non deceptive condition, both the child and a competitor sat on the window side of the box and the child again pressed the button corresponding to the empty side to receive the treat. In the Opponent-deceptive condition the child sat on the side of the box with the window, the competitor sat on the opposite side and the child had to point to the button corresponding to the empty side, so that the competitor pressed this button and the child received the treat from the other side. Russell et al. proposed three possible outcomes, each corresponding to a unique conclusion:

1.) Group differences typical to the task (i.e., 3-year-olds perform more poorly than 4-year-olds and children with autism perform more poorly than children without autism) are removed only in the no-opponent condition, thereby demonstrating that a limitation in understanding minds is central to the difficulty usually inherent in the task.
2.) Typical group differences remain across all conditions, thereby highlighting that an executive function deficit is central to task performance.
3.) Comparison groups perform similarly across all three conditions, demonstrating that something about mechanising the task made the task more accessible to children who typically do not succeed on the task.

Using both comparative group data and data in relation to first trial and total trial performance, Russell et al. found support for all three possible explanations. However, the strongest support was for the third outcome; that is, something about the mechanisation of the task made the task more accessible to young children and children with autism. Each child performed 20 trials of the same test. Data were collected for the first trial and total correct across all 20 trials. First trial performance can be considered an indicator of the child‟s conceptual understanding. Subsequent trials provided the child with an opportunity to learn from feedback (i.e., receiving or not receiving a treat) and adjust his/her strategy. Subsequent trials also measured the child‟s ability to maintain correct responding. Improvement across trials would have, therefore, indicated executive functioning through learning and adapting strategy. Decreasing performance across trials might be thought to have indicated poor executive functioning in relation to maintaining attention and focus.

Data from the typically developing children supported all three possible conclusions. That is, children in all groups found the deceptive tasks specifically difficult, mechanisation of the task made the task generally more accessible, and executive functioning deficits impaired younger children. When comparing the performance of typically developing 3 and 4-year-old children Russell et al. did not find a difference between the age groups on the first trial. Similar first trial performance across age groups demonstrated that 3-year-old children do not have a specific deficit in relation to their ability to create a false-belief when compared with 4-year-old children. There was, however, a difference between the age groups across the 20 trials. Russell et al. attributed this difference to more consistent performance among 4-year-olds. The authors postulated that this difference was due to more limited executive functioning in the 3-year-olds, which made consistent responding more difficult. Furthermore, Russell et al. found a significant difference in responding across the three types of opponent conditions. Specifically, this difference appeared to be in relation to more errors being made in the Opponent-deceptive condition, which proved most difficult for both age groups. This indicated that tasks requiring deception were especially difficult for young children.

Children with autism showed no specific difficulty in tasks involving deception when compared to children with moderate learning difficulties. On the first trial, children with autism performed as well as children without autism. This is evidence that children with autism do not have a specific deficit in relation to creating false belief, when compared to children matched for verbal ability who do not have autism. Overall, across the 20 trials children with autism performed more poorly than children without autism; although, this difference was the result of almost ceiling performance of the children without autism on the Opponent-non deceptive condition. While this was not discussed by Russell et al., a possible reason for the difference in performance is that the children in this group had, on average, higher verbal mental ages (m = 5.95 years) than the children in all other groups (range = 4.91 -5.40 years, m = 5.24). Despite the fact that the difference in verbal mental age was not statistically significant, Liss et al. (2001) argued that a relationship between executive functioning and verbal ability is underestimated in research that does not specifically control for language ability. Furthermore, no significant difference was found across the opponent conditions for the 20 trials, despite a significant difference for first trial performance; where both groups performed best on the no-opponent condition. It is, therefore, most probable that both executive functioning and the presence of an opponent mediated performance on the tasks equally for children with and without autism. Moreover, mechanisation of the task appeared to improve task accessibility for children with autism.

Russell et al. noted that deceptive performance of children in their study was better than deceptive performance of similar children in studies that utilised the standard windows procedure, in which the child has to tell a competitor where to look. As noted earlier, the authors suggested that this might have been the result of the mechanisation of the procedure which could have made this task more accessible to younger children and children with autism. While Russell and colleagues suggested that, because of procedural differences, it is not possible to compare their findings to those from other studies employing the standard windows task, it does appear that the automated task was cognitively more accessible to younger children and children with autism. The authors proposed that the difference in performance might be in relation to difficulty experienced by young children and children with autism to inhibit verbal responding. Russell et al. suggested that inhibiting a physical response requires less cognitive resources than inhibiting a verbal response. This proposition is, however, not supported by the previously discussed findings of Joseph and Tager-Flusberg (2004); where children appeared to perform better on a verbal inhibition task than they did on a structurally similar physical inhibition tasks. To determine conclusively what factors might contribute to performance on these tasks it would be necessary to compare performance on tasks for which only the element of pointing or verbalising differs.

In summary, the role of executive functions in performance on ToM tasks is still unclear. Mixed research findings and research limitations make drawing conclusions difficult. It is likely, however, that the relationship between performance on ToM tasks and executive functioning tasks is related to similar cognitive demands inherent to both types of tasks.

Language

It is not surprising that language has been reliably linked to ToM in typically developing children (Milligan, Astington, & Dack, 2007, for a comprehensive review) and ToM in children with autism (Happé, 1995, for a comprehensive review). Language is a central feature of human interaction. Needless to say, it plays a prominent role in typical child development, particularly so in the preschool years; which has also been remarked in research investigating the development of intelligence (Anderson, 1992). Furthermore, an absence of language and/or differences in acquired language ability is one of three main diagnostic features of autism (DSM-IV: American Psychiatric Association, 1994).

Astington and Baird (2005) presented an eloquent summary of the complex nature of the relationship between ToM and language, highlighting the mesh of interconnectivity between children‟s linguistic competence, children‟s communicative environment, and children‟s ToM. Research has approached the study of language and ToM in two ways. The first type of research examines the impact of children‟s linguistic competence on ToM performance (Astington & Jenkins, 1999; Fisher, Happé, & Dunn, 2005). This type of research typically measures or controls for linguistic ability, assessed by various types of developmental language measures, in relation to success or failure on tasks designed to measure ToM understanding (Fisher, et al., 2005; Happé, 1995; Milligan, et al., 2007; and many more). The second type of language-related ToM research investigates the impact of early communicative environment and experiences on later development of ToM (e.g., Brown, Donelan-McCall, & Dunn, 1996; Dunn, Brophy, Astington, & Baird, 2005; Farrant, Fletcher, & Maybery, 2006; Harris, Astington, & Baird, 2005; Jenkins & Astington, 1996). A summary of these two types of research follows.

Linguistic Competence and ToM

Due to the nature of ToM tasks, verbal ability has often been accounted for in ToM research. Most ToM tasks have substantial verbal components. Although the amount of language inherent in tasks varies substantially, generally speaking, children are told stories, asked questions, and sometimes required to respond verbally. Subsequently, children are either matched for verbal ability or the contribution of a target variable to ToM performance is measured after controlling for verbal ability (Happé, 1995; Milligan, et al., 2007). The breadth of research and volume of data generated in relation to language and ToM has resulted in two high impact reports; the first by Happé and a more recent meta-analysis by Milligan and colleagues. Where Happé examined the relation between language, age, and ToM in children with autism and typically developing children, Milligan, et al. examined the complexity of this relationship in typically developing children alone. Essentially, both studies highlighted the importance of language for ToM, and these findings will now be presented and evaluated.

Happé examined whether language ability contributed to ToM ability in children with autism and typically developing children. The study stemmed from a need to account for different findings in relation to ToM competence in children with autism, where the majority of studies in this field have found differing degrees of success on a range of ToM tasks (for a table summarising the range of findings see Happé, 1995, pp. 844-845). Happé utilised a larger than normal sample of data collected over a period of five years from a single university research unit, including data from 70 children with autism, 34 with an intellectual disability, and 70 typically developing children. Data existed for two standard ToM tasks (Sally-Anne task and Smarties task), the British Picture Vocabulary measure (both verbal IQ score (VIQ) and verbal mental age equivalent score (VMA)), and chronological age. Children who failed both tasks or passed only one task were collapsed into a single group, termed „fail".

Overall, Happé found a relationship between verbal ability and ToM in typically developing children and children with autism. Children with autism who passed both ToM tasks (n=56) had higher VMA (m = 9 years 7 months, SD = 46.4 months) and verbal IQ (m = 72.5, SD = 22.1) than those in the fail group (n=14; mean VMA = 5 years 5 months, SD = 21 months and mean VIQ = 48.9, SD = 14.8). Similarly, typically developing children who passed both tasks (n=39) had higher VIQs (m = 108.2, SD = 12) than those in the fail group (n=31; mean VIQ = 97.1, SD = 14.7). When divided by pass/fail on ToM tasks, no significant differences for any measure were found for children with an intellectual disability, although patterns of differences for VMA and VIQ were similar to those of the other two groups. Additionally, regression analyses indicated that only VMA predicted ToM in children with autism, whereas VMA, chronological age, and gender predicted ToM in typically developing children.

A later study brought together what was by now a much larger field of research to examine how performance on ToM tasks was related to different aspects of verbal ability. Milligan, et al. (2007), conducted a meta-analysis to examine the role of varying aspects of verbal ability for ToM, as assessed by a range of verbal measures employed in ToM research. The researchers aimed to examine which verbal abilities were most strongly related to ToM. To do this they compared the correlations found between ToM and the broad range of verbal measures used. Nineteen different tests were classified into one of five types of language measures: general language; semantics; receptive vocabulary; syntax; and memory for complements.

Milligan et al. reported that across studies language measures accounted for between 0 – 77% of performance on ToM measures; 0 – 40 % after controlling for age. On average, memory for complements measures were most strongly related to ToM measures and receptive vocabulary measures were least related, accounting for 44% and 12% of variance, respectively. Despite the range of variances, post hoc tests found that the only significant difference in strength of relationships was between receptive vocabulary and general language measures, where the latter accounted for 27% of variance in ToM ability. The authors proposed that, in part, this may be due to the more common use of these types of measures and, subsequently, their larger representation in the data set. The authors further proposed that a number of the types of measures overlap in the types of language ability they assess. That is, skills targeted in one test are required for success in another test, despite not being the focus of the other test. Milligan et al. explained the smaller association between vocabulary measures and ToM, by proposing that vocabulary measures were the only type of test able to isolate a single language ability, and therefore the most pure test of an isolated ability. The only other variable that explained the degree of variation in correlations between ToM and verbal ability across studies was the number of types of false-belief tasks used. Milligan et al. suggested this was to be expected, given that studies using multiple ToM tasks were likely to generate greater variance in ToM scores and, subsequently, stronger statistical relationships.

Milligan et al.‟s findings clearly demonstrated the importance of language for ToM. While language measures differed in both structure and target skill, they were almost equally, statistically speaking, related to ToM. Given the imbalance of studies using less common language measures, it might be reasonable to conclude that the most strongly related measure was that which assessed the most ToM specific language structure; memory for complements. Memory for complements is, therefore, likely to be most strongly related to ToM because an understanding of complements is necessary for ToM understanding or ToM competence.

Memory for complements was likely most strongly related to ToM because it is reflective of linguistic structures imbedded in most ToM tasks. While others have treated children‟s verbal ability as a mediating factor, which enables children to demonstrate their understanding of mind (e.g., Happé, 1995; Lewis & Osborne, 1990), de Villiers, proposed that acquisition of linguistic structures plays a much more important role in ToM development (for a comprehensive summary see de Villiers, J. G., Astington, J. W., & Baird, J. A., 2005; de Villiers, P. A., Astington, J. W., & Baird, J. A., 2005). This perspective proposes that grammatical complement structures act as a scaffold for later learning about representations. Jill de Villiers and colleagues claimed that children demonstrate difficulties with non-mental state questions that have a grammatical structure similar to that of ToM tasks. The following is an example:

The Mom said she bought apples, but look, she really bought oranges.

What did the Mom say she bought? (p.188)

According to de Villiers et al., 3-year-olds generally respond with "oranges" but 4-year-old children generally respond correctly with "apples". Both the verb „said‟ and mental state verbs such as *think* and *believe* require more complex grammatical structures than mental state verbs such as *want*. For example you cannot say "I think an apple", but instead you might say "I think that is an apple". Clearly children need to have a functional grasp of these grammatical structures, termed embedded complements, before they can talk or think about other people"s beliefs or thoughts. Proficiency in using complement structures is, however, not sufficient for a fully developed understanding of mind. Children must also have acquired knowledge of how people come to think/believe and how these mental states can impact behaviour. Despite the relative intuitive and common sense nature of the linguistic determinist position adopted by de Villiers, the theory cannot account for a number of empirical findings of the ToM research domain; these are highlighted by de Villiers.

The most problematic criticism was made by Perner and colleagues (Perner, Zauner, Sprung, Astington, & Baird, 2005). They argued that ToM research from non-English speaking backgrounds has found that children typically understand *think* later than *want,* even when these two mental state verbs are used in grammatically identical ways. Here, the differential acquisition of the verbs, *want* and *think*, demonstrates children"s acquisition of conceptual understanding rather than linguistic understanding. Perhaps Milligan et al."s results best demonstrated a shared contribution, from both environmental factors and linguistic artefacts. It can be argued that the range of relationships across language ability measures demonstrates the impact of environment, where shared environmental factors contribute to both broad language development and ToM. Additionally, Milligan"s data can be used to demonstrate that the relationship between ToM and language is strongest when both environmental and linguistic contributions are assessed, as is the case when memory for complements measures are utilised. It would be reasonable to conclude that early communicative environment contributes to both ToM and memory for complements; for this reason a degree of similarity is expected. Additional similarity, that is similarity above that found in relation to other language measures as found by Milligan et al., might then be due to shared linguistic artefacts.

To summarise, ToM competence most likely relies on a combination of linguistic maturity and conceptual understanding. The way in which these fundamental components are acquired is, arguably, an important topic for research, and an avenue of research that is likely to contribute to the development of effective interventions for children who do not develop neuro-typically. A likely avenue for the acquisition of these concepts is through experience.

Early Communicative Environment and ToM

Children‟s early communicative environments are composed of a number of factors. Examples include family size, order of birth, parental socio-economic status, care arrangements, parental disciplinary beliefs, type and range of talk between parent and child, and so forth. The role of early environmental factors in the development of ToM has been widely examined and a brief overview of the key findings will now be presented.

Twin studies have generated mixed findings about the role of environmental and genetic factors for ToM. Research using a relatively small sample (N = 119 pairs of 3-year-old same sex twins) found genetic factors to be most important for early ToM (Hughes & Cutting, 1999), generally, therefore, supporting an innate account of ToM. However, later research, using a much larger sample (N = 1116 pairs of 5-year-old same sex twins) found that environmental influences contributed most to ToM (Hughes, et al., 2005). Hughes, et al. (2005) suggested that their later study was better able to detect shared environmental influences because it had a much larger and more appropriate sample size for the required analyses. Having sampled from a much broader socio-economic range, Hughes et al. argued that broader environmental conditions were detected in the later study. Age of the samples might also have contributed to differences in findings. That is, at the age of 3 years twins are likely to have more homogenous environments compared to twins at age 5 years. Hughes acknowledged that by the age of 5, children have generally had peer-experiences outside the home; experiences which can contribute to the development of ToM. Examples of the types of environmental influences for ToM follow.

Early relationships influence the types of early experiences that children have, and consequently how children develop. For example, research has reported that older siblings have a positive effect on a child‟s development of ToM (Jenkins & Astington, 1996; Perner, Ruffman, & Leekam, 1994). This effect is likely the result of the types of language interactions between children and their peers; interactions that differ from those commonly observed between children and their parents (Brown, et al., 1996). Brown, et al., found that children typically have more conversations about mental states with siblings and friends than with their mothers. However, various studies have found that mother-child conversation is important for the development of ToM (Brown, et al., 1996; de Rosnay & Hughes, 2006; Dunn, et al., 2005; Meins & Fernyhough, 1999; Ontai & Thompson, 2008; Slaughter, Peterson, & Mackintosh, 2007; Taumoepeau & Ruffman, 2008). Overall, research has indicated that both type of conversation and frequency of conversation are predictive factors for ToM in typically developing children.

Atypical early experiences further highlight the importance of early language environment. Autism is the most commonly used example of a population that experiences ToM delay or deficit, as has been discussed earlier in this chapter and in Chapter 1 of this thesis. The cause of ToM differences in autism is hotly debated. One explanation is that children with autism are not readily involved in early interaction and shared experiences, which are necessary for the development of an understanding of other‟s minds (Hobson, 2004). Similar delays in ToM have been found in populations of children who are born deaf, and whose parents are not fluent in sign language early in the child‟s life (de Villiers, P. A., et al., 2005; Peterson & Siegal, 2000; Peterson, 2004; Peterson & Siegal, 1995; Peterson & Wellman, 2009; Peterson, Wellman, & Liu, 2005). This research has found that early linguistic experiences are most important for the development of ToM (Peterson, 2004). Early communicative experience and, in turn, ToM development of late signing children is vastly different to that in those children who are born deaf to deaf parents and therefore exposed to early fluent sign language (for a review see Peterson & Siegal, 2000). While the underlying cause for early social disturbances are no doubt different for children born deaf and children with autism, it is clear that atypical early experiences can result in serious consequences for later social development.

To summarise, research has demonstrated the key role played by early social experiences in the development of ToM. Furthermore, an absence of typical experiences can lead to later social and ToM difficulties. ToM delays in children with deafness and autism demonstrate the vital role of children‟s early social environments. The cause of the disruption in children born deaf is obvious; children do not experience typical language interactions due to a parent‟s inability to communicate effectively with their deaf children. On the other hand, the cause of early social environmental disruption in autism is not entirely clear. It is plausible that a non-social disruption, which has consequences for social interaction, could cause ToM deficits in autism. Some of the symptoms of autism might be the result of damage to a specifically social mental module. It is, however, equally likely that an inability to engage with the social world due to sensory processing difficulties results in later social difficulties (Williams, 2004).

Research Aims

ToM research appears to have moved on, in part, from some of the important issues outlined in this chapter, to some extent suggesting these issues have been resolved. In the opinion of this reviewer, however, a number of issues remain unresolved. Certainly, this thesis does not attempt to resolve them all. The focus of this thesis will be the connection between ToM, cognition and language. Particularly, studies within this thesis will address the developmental and cognitive profile of children who succeed on ToM tasks. The literature has reported mixed and varying findings that both cognitive and language factors that are important for ToM in both young children and children with autism. This thesis, therefore, aims to investigate differences, if any, in the broad skill sets that underpin ToM ability in both typically developing children and children with autism. Examination of how these skills are acquired is, however, beyond the scope of this thesis. For that reason, early social environments will not be measured. Instead current cognitive profiles will be assessed.

CHAPTER 3 - Pilot Study of the Relationship between ToM Task Success and Developmental Ability in Children With and Without Autism.

The literature reviewed in the previous chapter reported mixed findings about the relationships between ToM and other aspects of child development (i.e., executive functions, reasoning, and language). In the previous chapters it was proposed that mixed findings were likely to stem from the use of different tasks and procedures across studies; for both ToM tasks and non-ToM tasks (e.g., tasks associated with factors thought to be related to ToM). Some support for this claim comes from large meta-analyses that have found that different measures give different results (Milligan, Astington, & Dack, 2007; Wellman, Cross, & Watson, 2001). These meta-analyses have found that not all types of differences across studies contribute to differences in ToM outcomes. Instead, a few isolated factors have been identified. A brief recapitulation of these factors follows.

Milligan et al. (2007) found that differences in the degree of association between language and ToM were related to the type of language measure used. Specifically, the only significant difference was between the most commonly used language measures (general language ability measure and specific vocabulary measures). Other language measures that targeted specific understanding (e.g., understanding of complements) had only been used in a few studies; hence power to detect differences was limited. Aside from differences in the measures used to assess the role of language, measures assessing ToM have also differed in a number of ways. Wellman et al. (2001) assessed the effect of 13 types of variations across studies and found that seven of these impacted on ToM scores. Briefly, these were: age; motive of the transformation in the ToM task; child‟s participation in the task; real presence of the target object; salience of the protagonist‟s mental state; child‟s country of origin; and use of a temporal marker in the test question. Those that were not found to contribute to ToM were: year of publication; type of ToM task, type of test question; nature of the protagonist; nature of the target object; and self versus other‟s ToM.

In short, task and procedural differences can result in mixed findings across studies. In turn, this can cause uncertainty and, importantly, doubts about the nature of what ToM tasks measure. Certainly, for this researcher, differences in findings across the broad spectrum of research have raised questions about what precisely is measured by ToM tasks. Moreover, it has raised questions about the nature of the web of relationships between ToM and global development in children with and without autism.

Although there is clearly a difference in ToM between children with and without autism, how this difference relates to broad abilities in each of these groups remains unclear. In part this is because methodological differences across studies and differences in ToM and non-ToM tasks used by studies confound findings. There is limited evidence that the order of acquisition of mental state concepts differs in children with autism (Peterson, Wellman, & Liu, 2005). Tager-Flusberg and Joseph (2005) argued that different routes lead to ToM competence in typically developing children and children with autism. They proposed that two main components underpin ToM in typically developing children; namely social-perceptual and social-cognitive components. In other words, ToM involves, firstly, orienting to and processing social stimuli; and secondly, the capacity to reason about social phenomena. These authors suggested that children with autism use language ability to compensate for their social-perceptual deficit. Success on ToM tasks in this population ought, therefore, to be more dependent on language than is so in children without autism.

In order to examine whether different abilities contribute to ToM understanding, the current pilot study re-analysed data collected by the author for an honours thesis project. This pilot study aimed to assess the feasibility of using a developmental measure to study the foundations of ToM in children with and without autism. The original data collected included a measure of ToM along with developmental measures of language, cognition, and personal-social development. Although useful, the current pilot study was not able to evaluate the contribution to ToM from the gamut of factors hypothesised to be important for ToM. Instead the study looked to assess the practicality of using a broad developmental measure to assess early childhood abilities, in children with and without autism, in relation to ToM. While not all of the abilities assessed by the developmental measure have been implicated in ToM – and nor would we expect them to be - broad measurement of abilities in all of these areas, will highlight differences between children with and without autism beyond those associated with ToM. It is proposed that this could generate developmental profiles of children with and without autism who successfully develop ToM and those who do not. While these profiles might differ in ways that are related to ToM, children may also differ in non-ToM related abilities. These differences might be thought to demonstrate underlying causes of both ToM and non-ToM deficits.

Method

This method has been previously presented in the honours thesis (2005) of the current author. A simplified version, encompassing the information relevant for this pilot study is presented below.

Participants

Twenty children participated in this study, 10 with autism (1 female, 9 males) and 10 children without autism (3 females, 7 males). The original data were collected in order to compare ToM outcomes for children with autism who had a history of intensive early one-on-one intervention to children with autism who had did not have a history of intensive early one-on-one intervention. For the purpose of the new analyses children with autism were treated as a single group. Sample characteristics are presented in Table 1.

Only children with a verbal age above 4-years old (as measured by the Communication Domain of the BDI) were included in the analyses. Children recruited for the autism group were only included if they had a current diagnosis of an autism spectrum disorder from a qualified psychologist. Children in the typically developing group were only included if they had no history of a developmental or behavioural disorder.

Children were recruited through either: an advertisement in a local autism association newsletter; through public schools; or through community groups familiar to the researcher. Most participants were obtained through schools on behalf of the researcher and through community groups known to the researcher. One participant was sourced through an advertisement in the local autism association newsletter.

Materials

The Personal-Social, Communication, and Cognitive Domains of the Battelle Developmental Inventory (Newborg, Stock, Wnek, Guidubaldi, & Svinicki, 1984) were administered to measure developmental maturity in areas hypothesised to be linked with ToM. The three domains were administered according to the instructions for the BDI.

The BDI was selected because it is to able assess splintered development. It has been found to have good reliability and validity for both typically developing children and children with developmental disabilities (Glascoe, Martin, & Humphrey, 1990). Each of the BDI domains assess a number of areas of development; from birth to 8-years-old. The Personal-Social Domain measures children‟s concept of self and social role, ability to relate to adults, and ability to relate to peers. The Communication Domain is a broad language measure divided into two components; receptive and expressive language ability. The Cognitive Domain is a broad abilities measure divided into three components: attention and memory; perception and concepts; reasoning and academic skills. BDI produces raw scores, age equivalent scores, and scaled scores for each domain. Age equivalent scores were used here, because they provide a better means of comparing specific abilities across heterogeneous groups.

Table 1

Sample Characteristics

	Typically Developing n = 10, M (SD)	Autism – Intensive Intervention n = 5, M (SD)	Autism – No Intensive Intervention n = 5, M (SD)
Chronological Age (Months)	70.30 (12.63)	95.60 (31.77)	116 (40.46)
BDI Personal-Social age in months	77.35 (10.21)	70.10 (15.54)	53.20 (9.52)
BDI Communication age in months	78.40 (12.50)	75.10 (15.22)	56.90 (17.41)
BDI Cognitive age in months	75.70 (10.15)	77.40 (10.26)	62.10 (15.33)
Number passing ToM task	7	4	3

To measure Theory of Mind ability the researcher administered the Sally-Anne false-belief test as described by Baron-Cohen et al. (1985) using props; dolls (Ted and Minnie) and cardboard boxes that differed in colour, shape and size (small red oval and big purple rectangle). Control questions (Did Minnie see Ted move the ball?; Where is the ball now?; Where did Minnie put the ball?) were asked after the ToM question ("Where will Minnie look for the ball?).

A standard consent form (for the participation of children in research, and those with a disability) was used to obtain consent from parents or carers of the participating children. A letter outlining what was involved in participating in the study, risks associated with participation, and the purpose of the study was given to parents, to enable informed consent. A similar letter requested the involvement of school principals in sourcing participants.

Procedure

Children were tested in their homes, schools, and in two instances in a testing room on the premises of the School of Psychology at the University of Adelaide, where this research project was based. In all instances where the child was tested at home or at the School of Psychology the parent was present and the researcher explained the procedure of the study to the parent. Parents were provided with the information letter and the consent form prior to commencing testing. In instances where the child was tested at their school, the parent had been previously sent the relevant information and consent forms and the researcher collected the consent forms from the school principal.

The order of developmental test administration was as per the instructions in the BDI handbook; Communication Domain followed by the Cognitive Domain. These were followed by the Sally-Anne story, the false belief test question and the control questions, in that order. Order of false belief test question and control questions has varied from study to study, but this researcher believes control questions might act as a prompt for correct responding for some children and not for others, and therefore these were asked after the test question. The Personal-Social Domain was administered last, because this required interviewing the parent or teacher of the child.

All children who participated completed the entire test battery. Testing took on average two hours per child (range 40 mins – 3.5 hours) and was for the most part administered in a single session. In two instances the participating child, once a child with autism and once a child without autism, became tired and needed to be assessed over two sessions. Both children were younger participants and, subsequently, testing took longer with these children. Splitting testing across sessions helped to insure children were focused throughout. Testing across multiple sessions is allowable, based on the BDI handbook instructions. In both of these instances, the Communication Domain was administered on the first visit, and the Cognitive Domain and false-belief test were administered on the second visit.

Results

To examine whether a diagnosis of autism impacted on the relationship between ToM and development, children in each diagnostic group were further divided into two groups based on ToM ability; that is children who passed (passers) and children who failed (failers) the ToM task. Figure 1 and Table 2 illustrate a clear pattern of results. That is, both children with autism and typically developing children displayed similar developmental patterns when divided into those who passed the ToM task and those who did not. In brief, children who passed the ToM task had better communication skills, were more cognitively developed, and were more socially developed. However, the small sample size limited power to detect differences between passers and failers. For children with autism the only significant difference between abilities between passers and failers was for communication age. For typically developing children there were significant differences between passers and failers for both communication age and chronological age.

Given the small sample size, it is useful to examine effect sizes to determine the size of potential differences between the groups. As expected, the largest effect sizes were in relation to communication scores for both diagnostic groups; but these were largest in the typically developing group. Small to moderate effect sizes were found in relation to cognitive abilities in both diagnostic groups. Personal-social development showed the smallest effect size, again in both diagnostic groups.

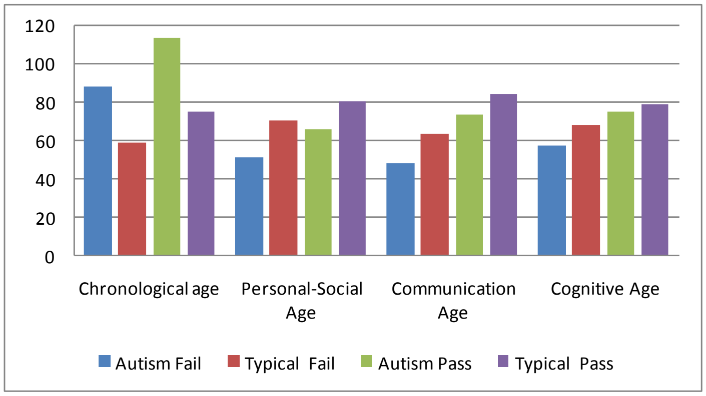

Figure 1. Pattern of developmental abilities; differentiated by diagnosis and ToM ability; age in months

Table 2

Comparison of Passers and Failers on Chronological Age and Domains of the BDI

	Autism				Typically Developing			
	ToM – Yes (n=7) M (SD)	ToM – No (n=3) M (SD)	*p*	partial eta^2	ToM – Yes (n=7) M (SD)	ToM – No (n=3) M (SD)	*p*	partial eta^2
Chron. Age (months)	113.29 (36.03)	88.33 (35.57)	.34	.11	75.29 (10.19)	58.67 (10.79)	.048	.40
P-SI Age (months)	66.21 (14.60)	51.00 (11.79)	.15	.24	80.21 (9.99)	70.67 (8.55)	.19	.20
Com. Age (months)	73.71 (16.13)	48.00 (3)	.02	.47	84.71 (6.76)	63.67 (10.07)	.004	.66
Cog. Age, (months)	75.07 (12.19)	57.33 (14.05)	.07	.34	78.93 (8.22)	68.17 (11.79)	.13	.26

Note: Chron. Age – Chronological age; P-SI Age – Personal-social age; Com. Age – Communication age; Cog. Age – Cognitive age.

Discussion

The findings from this pilot study provide a starting point for comparison of the development of ToM in children with and without autism. First and foremost, the data demonstrated the developmental nature of ToM in this sample of children. There was a clear pattern across the developmental domains, where children who passed the ToM test performed better than children who did not pass the test; with the exception of the personal-social domain, where children with autism who passed the ToM task scored lower than typically developing children who failed the ToM task. In light of the pattern of results, ToM in the current sample of children with autism ought to be considered delayed rather than deficient.

It would, however, be remiss to ignore the differences between the current findings and those more commonly, but not always, reported in the literature. While an approximately even pattern of development in typically developing children is not unusual, such steady development is not typical in autism. Children with autism typically show splintered development and isolated abilities (DSM-IV; American Psychiatric Association, 1994). Furthermore, there is a general consensus in the broad ToM literature, that individuals with autism achieve ToM success through processes that differ from those implicated in typically developed ToM. In this respect, the data from this study appear to be contradictory.

Specifically, the literature discusses success on tests of ToM in autism as being achieved through processes that compensate for a lack of automatic processing (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Leslie, & Frith, 1985; Colle, Baron-Cohen, & Hill, 2007; Fisher, Happé, & Dunn, 2005; Happé, 1995; Leslie, 1994; Leslie & Thaiss, 1992). Support for the theory, of different ToM processes, also comes from biographical accounts of what social interactions are like for people with autism (Williams, 2004). Additionally, the „different ToM in autism" position is to some extent supported by findings that the order in which children acquire understanding of mental concepts differs to some degree in children with autism

(Peterson, et al., 2005; Wellman & Liu, 2004). An examination of possible explanations for the apparent contradiction of the current data set follows.

One possible explanation, for better that expected performance in the group of children with autism, is that something about the ToM task made this task more accessible or more susceptible to guessing the correct answer. However, the task did not differ in any meaningful way from the tasks used in other research (Baron-Cohen, et al., 1985; Wellman, et al., 2001). It is therefore unlikely that the ToM task was unusually easy for children with autism.

A more plausible explanation for the current outcome is that there was something unique about the current sample of children with autism. The sample was recruited for a study that compared ToM outcomes in children who had a history of intensive one-on-one intervention and children who had had limited intervention. Interventions included music therapy, speech therapy, occupational therapy, applied behavioural analyses, play therapy, and picture exchange communication training. The only requirement for inclusion in the intervention group was that the child had received at minimum two years of 14 hours a week, one-on-one intervention. It is possible that one-on-one interventions helped promote the development of ToM because these interventions forced children with autism to interact with another person, which they might not otherwise have done. This may also explain why the children in this sample developed ToM sooner than is reported in the literature (Happé, 1995). The group of children with autism who passed the ToM measure were, however, made up of both children who had and those who had not received intensive intervention. Therefore, while intervention may account for some of the ToM advances made in this sample, it cannot explain all ToM ability of the children with autism.

Unfortunately, the sample from the current author‟s honours study was too small to determine whether there were any differences in ToM and development between children who had received intensive intervention and those who had not. It is still entirely likely that early intensive one-on-one intervention had positive consequence for ToM, even though interventions did not specifically target ToM related abilities. It is important to note that many interventions for children with autism are language based, and some might argue that the language aspect rather than the one-on-one aspect of the intervention is responsible for ToM gains, given what we know about the importance of language for ToM. The current study cannot differentiate between these aspects of intervention, but this might be a useful investigation for future research. Findings about the aspects of intervention which are important for ToM would contribute the understanding of the role of language and social environment in the development of ToM. For example, if differences in ToM were found to exist in children who have received various types of intensive one-on-one interventions, irrespective of their language components, it would provide support for the theory that ToM does not develop in autism because children experience limited social interactions. This would fit well with Carpendale and Lewis‟s theory of triadic interaction (2004).

Aside from the unusual ToM success rate of children with autism, the homogenous development across BDI domains also requires some explanation. The BDI was designed to test developmental strengths and weaknesses in children in order to best tailor interventions; and by all accounts it does this well (Glascoe, et al., 1990; Newborg, et al., 1984). The current finding, of similarity in age equivalent scores across domains, in the children with autism, is not entirely abnormal. To begin with, it should be noted that children with autism did have somewhat lower personal-social age equivalent scores than communication and cognitive scores; a difference that was not apparent in the typically developing sample. This fits well with autistic diagnostic criteria of social deficits (DSM-IV; APA 1994). Furthermore, similarities between cognitive age equivalent scores and communication scores might not be thought of as unusual. The cognitive domain relies heavily on language; in the form of verbal instructions for children and in some cases children‟s verbal responses. Therefore, similarities in scores across these domains might be the result of language ability mediating performance on both domains. Thought of in this way, homogenous performance in this sample across domains in children with autism is not out of the ordinary.

Finally, the relationship between ToM and abilities measured by some aspects of the BDI might be due to similar cognitive or executive demands that mediated performance on both types of measures. Certainly, weaker associations between ToM ability and personal-social development might be the result of no shared performance demands for these tasks. In comparison to the ToM task, the personal-social domain did not rely on executive functions because, for the most part, scores were calculated from parent/teacher interview responses rather than child responses. In light of possible shared variance in performance across all tasks, the degree to which ToM was related to communication ability and cognitive ability cannot be accurately assessed from the current data set.

Nevertheless, the results of the current study provide insight into the levels of varying skills that might be required to pass tests of false belief. Not surprisingly, effect sizes indicated that the strongest association was between ToM and language; in line with the robust relationship reported in the literature (Astington & Baird, 2005; Happé, 1995). Primarily, findings of the study indicated that overall the BDI was useful for comparing the abilities of children with autism to those of typically developing children in relation to ToM. However, to better differentiate between cognitive and language abilities, it would be useful to include an additional early abilities measure, such as an IQ test. Further investigation of ToM ability in relation to developmental ability but also IQ might help to isolate language and non-language differences between children who pass and fail ToM tasks in cognitive areas. Additionally, a follow up study should include a measure of reasoning that does not draw on mental concepts, but that is similar to ToM tasks. Inclusion of these could help differentiate between links with developmental measures and ToM that are specific to mental reasoning, rather than generally linked to reasoning or executive functions. Finally, to ensure reliability of ToM scores, a number of ToM tasks should be included.

CHAPTER 4 - Correlates of Theory of Mind Change with Age in Normally Developing Children: The Role of Language and Reasoning.

Abstract

To examine the link between ToM and cognitive development, 60 typically developing children (30 4-year-olds and 30 6-year-olds) each completed four ToM measures of varying difficulty and several cognitive measures; two subtractive reasoning measures, the Concept Formation and Verbal Comprehension subsets from the Woodcock Johnson III measure of intellectual functioning and the Cognitive and Communication Domains from the Battelle Developmental Inventory 2 (BDI-2). Correlates of ToM performance were not stable across age groups. Language skills, as measured by the BDI-2, correlated with ToM in 4-year-olds. In 6-year-olds, however, only subtractive reasoning related was related to ToM performance. The role of language in the development of ToM is crucial, but it appears that, as children mature, ToM becomes less dependent on language ability. Instead, reasoning competence begins to differentiate children who pass or fail ToM tests.

Correlates of Theory of Mind Change with Age in Normally Developing Children: The Role of Language and Reasoning.

Premack and Woodruff (1978) termed the cognitive processes that enable people to predict and explain the behaviour of others "Theory of Mind" (ToM). This capacity relies on the fundamental understanding that other people have thoughts and beliefs that may differ from one‟s own thoughts and beliefs. A false-belief (FB) test (Wimmer & Perner, 1983) has been widely used as a standard measure of ToM (Astington, 2003; Baron-Cohen, Leslie, & Frith, 1985; Fisher, Happé, & Dunn, 2005; Jarrold, Butler, Cottington, & Jimenez, 2000; Joseph & Tager Flusberg, 2004; Peterson, 2004; Sabbagh & Taylor, 2000) because it targets a person‟s capability to take the position of another person, or more specifically, it assesses what one person judges another person to know. Success in this test requires the individual to use his/her understanding of the knowledge of the character in a scenario to predict behaviour consistent with that character‟s belief, rather than with the individual‟s own knowledge, which differs from that available to the character in the scenario.

The emergence of ToM is linked to early childhood development. Thus, in a series of experiments, Wimmer and Perner (1983) found that 3-year-old children did not pass FB tests, but just over half of all children between the ages of 4 and 6 years passed, and most aged between 6 and 9 years passed. Not surprisingly, since then children have been the focus of this area of research. By dividing children into groups according to age and determining which groups perform above chance on FB tests, research has established that by age 4, approximately 50% of children are able to answer FB questions correctly (for reviews see Dissanayake & Macintosh, 2003; Flavell & Miller, 1998; Lewis & Carpendale, 2002; Slaughter & Repacholi, 2003). However, Slaughter and Repacholi (2003) emphasised the limitations of defining acquisition of ToM in terms of groups that perform at or below chance on FB tests or test batteries, noting that this leads to an all-or-nothing view of ToM. Research suggests that this is, indeed, not representative of the nature of ToM, which is likely to emerge over time (Slaughter & Repacholi, 2003).

A number of theories exist about the nature of ToM. Many of these theories attempt to explain one of the early ToM research findings, that people with autism lack ToM. Baron-Cohen et al. (1985) first proposed that people with autism lacked ToM and that this was an underlying cause of their social difficulties. Later, Charman and Baron-Cohen (1992) found that most children with autism exhibited a delay in ToM greater than that seen in children with other intellectual impairments, as well as that for typically developing children. Since then, a great deal of research has attempted to find alternative explanations for the social deficiencies that accompany autism. Some researchers have postulated, following Fodor (1983), that there is an all-or-none input system (termed "module" by Fodor) that mediates social functioning and that is missing in individuals with autism (Leslie, 1994). An account of how this module emerges, operates, and where it is located is still being formed (Friedman & Leslie, 2004; Leslie, German, & Polizzi, 2005). Others, however, have rejected such a reflexive account, arguing that ToM is an redundant conecpt, proposing instead that shared and co-regulated emotional experiences are the critical factors that promote a child"s learning about the world and that these experiences are limited, for to date unknown reasons, in individuals with autism (Shanker, 2004).

Although many theories have tried to explain why ToM is often absent in autism and numerous studies have examined ToM before the age of 4 years, very few studies have examined ToM in older children and adults without any intellectual disability (Friedman & Leslie, 2004; Slaughter & Repacholi, 2003). While the emergence of ToM in young children is interesting, a lack of research in older children and adults is problematic for theories about ToM, because ToM sophistication continues to develop well beyond its initial emergence in young children. This is evidenced in isolated studies that have found better performance on more complex ToM tasks in older children (Flavell & Miller, 1998).

A pattern of increasing performance in ToM tasks is hypothesised to be due to increased cognitive abilities (such as memory, attention, reasoning, and language) that may allow children to demonstrate ToM ability. What is arguably needed for ToM reasoning to occur is, firstly, the understanding that others have thoughts and beliefs that differ from one‟s own and, equally importantly, the cognitive ability to reason and communicate about that knowledge. It is useful then to conceptualise ToM as a skill reliant on prior acquisition of subskills. Wellman, Cross, and Watson (2001) made this distinction between competence (the knowledge and understanding that other people have thoughts, beliefs and desires that can differ from ones‟ own, that are required to solve the problem) and performance (other cognitive skills like memory, attention, and comprehension that are required to follow and remember aspects of the FB test stories). They noted that both are necessary to pass FB tests.

In a summary of research examining individual differences in ToM, Slaughter and Repacholi (2003) drew attention to significant positive correlations between ToM and a range of cognitive constructs such as executive functioning, fantasy, creativity, moral reasoning, and language. Nonetheless, despite the large number of positive correlations between ToM and several cognitive constructs, ToM and cognitive development (in such areas as executive functioning and IQ) have not been found to be causally related, and only language development has been reliably found to predict ToM performance (Astington & Jenkins, 1999).

A recent meta-analysis of the nature of the association between language and ToM found variation across effect sizes for the type of language measure used; the largest effect size was for the relation between ToM and memory for complements, followed by syntax, general language measures, semantics, and finally receptive vocabulary measures (Milligan, Astington, & Dack, 2007). Despite a main effect for type of language measure, Milligan et al. (2007) reported that a significant difference between mean effect sizes exists only between general language measures and vocabulary measures. These authors proposed that receptive language measures are the least strongly related to ToM measures because they target an isolated ability and do not overlap with other abilities, as is the case for the other language measures assessed. It is, however, also likely that the receptive measures are least related, although still accounting for 12% variability, because they are least sensitive to developmental variations among children. A general language measure, for instance, is better able to assess broader functioning, and is perhaps more representative of a child"s development in preschool age cohorts. Given that responding to ToM tasks involves a range of abilities, for which competency can vary in young children, it is likely that a general measure will better capture developmental variations that may also impact ToM development. Furthermore, the meta-analysis used data from children under the age of 7 years, but did not specify mean age or exact age range, which reduced sensitivity of the analysis to variation in relations between the constructs at the age when ToM begins to emerge. The authors also highlighted this issue, reporting that "maturational factors" may have affected variability of effect sizes across studies (Milligan, et al., 2007, p. 638).

The consistent finding that ToM is linked to verbal ability in children with autism (Happé, 1995) also suggests that a minimum level of language is necessary before children can respond successfully to ToM tasks. Given the large number of correlations between ToM and cognitive abilities, it is likely that cognitive skills underpinning performance on ToM measures develop on a similar age trajectory to ToM. It follows that, even if ToM were to be confirmed as an automated process located in a mental module, using ToM successfully still depends on language and reasoning skills.

One method of decoupling the mind-reading component (i.e., the ability to infer belief of another person) from any other cognitive components (e.g., language, memory, and attention) involved in answering FB tests is to use the same tasks with the same cognitive demands but to include probe questions that do not involve reference to another‟s thoughts. In this way, Riggs, Peterson and Bowler (2000) identified one aspect of the FB task to be counterfactual reasoning; that is the correct answer to the test question conflicts with the known reality of the person being asked the FB question. They further proposed that a reasoning strategy that they termed subtractive reasoning (SR) embodies the counterfactual or hypothetical aspect of FB reasoning. Their study found that SR skill is a necessary but not sufficient component of successful FB test completion.

The SR test is equivalent to the FB test in every aspect except that the test question, is directed to a different referent. Instead of asking the child to predict the action of a character based on that character‟s belief, the SR test question asks the child to reason hypothetically about the location of an object, assuming that the event whereby the object was moved did not take place. Arguably, the FB question requires further cognitive processing than the SR question, but Peterson and Bowler (2000) proposed that generating the reasoning involved in answering the SR question is a required prior step in the process of answering the FB question. Table 1 illustrates the FB and SR questions, along with a potential reasoning process for each. The FB task requires the type of hypothetical reasoning „if not …then…‟ also needed to answer SR questions. This approach is believed to isolate the mind reading component of FB tests. Children who pass SR tests, which do not require reference to the mind, but do not pass the FB tests, would therefore be deemed specifically to have trouble referencing another‟s mind.

If ToM is linked to age, it is necessary to locate it within cognitive development. Studies examining the relationship of ToM to cognitive abilities have implicated verbal ability, and in some cases IQ. In this study we examined more closely the relationship between ToM, SR, intellectual ability, and other abilities for memory, attention and language use, in children who were developing early ToM and children whose ToM was further developed. We considered that language is significantly advanced by 6 years of age and therefore other markers of cognitive development are likely to become increasingly important for children"s ToM. We tested the hypotheses that: ToM test performance is positively related to language competence, intellectual ability, and subtractive reasoning ability; language predicts ToM performance in 4 year old children; and that predictors of ToM performance are not stable between the ages of 4 and 6 years.

Table 1

Proposed Reasoning Steps Involved in Answering FB and SR Test Questions.

False-Belief	Subtractive Reasoning
Where will Paul look for the chocolate?	*Where would the chocolate be if Mary had not baked the cake?*
Paul was gone when Mary baked the cake.	When Mary baked the cake she moved the chocolate.
Therefore, Paul does not know that the chocolate was moved.	**If she had not** baked the cake **then** the chocolate would be where it was before she baked the cake.
If he does not know where the chocolate is now, **then** he will think it is still where he put it.	Before she baked the cake, the chocolate was in the cupboard.
He put it in the cupboard.	
So he will look in the cupboard.	

Method

Participants

Sixty children participated, 30 4-year-olds (age M = 4 yrs:8 mths, SD = 3 mths, range 4:1 – 5:1 yrs, 18 girls, 12 boys) and 30 6-year-olds (age M = 6:6 yrs, SD = 6 mths, range 5:4 – 7:7, 13 girls, 17 boys). Recruitment was through preschools, schools or through community groups familiar to the first author. The only exclusion criterion was having any identified intellectual disability or developmental disorder. Consent for participation was obtained from parents or carers and preschool directors or school principals. Children were from middle to high socioeconomic groups, all attending preschool or school in metropolitan public or private schools in Adelaide, South Australia.

Materials

Language and Cognitive Development

The Communication and Cognitive domains of the Battelle Developmental Inventory-2 (BDI-2) measure developmental maturity (Newborg, 2005). This measure was used in addition to a standardised test of IQ, used in much of the published research, because it differentiates between more specific aspects of cognitive development that cannot be directly measured by IQ tests alone, in children aged 0-8 years. Furthermore, the BDI-2 was not used by schools involved in the research project for any assessment related to academic performance and it therefore was not vulnerable to test-retest problems that might arise if children had been assessed by the school or education department or needed to be assessed in the future. The Communication domain of the BDI-2 measures general language in the areas of receptive and expressive language found to be reliably related to ToM (Milligan, et al., 2007); both are important factors in comprehension and response to ToM stories. The Cognitive domain measures developmental achievement in three areas of cognition: attention and memory capacity, perceptual and conceptual skills, and reasoning and academic skills. Attention and memory may be necessary components for FB story comprehension. Perceptual and conceptual development measures how much a child has learned about the physical world (e.g., an item that measures whether the child has acquired the Piagetian construct of object permanence and another that measures how well a child can group objects into categories). Testing perceptual and conceptual development may also indicate how much a child has also learned about people (e.g., if people do not see something happen, they do not know that it has happened). Reasoning and academic skills are measured in relation to children‟s acquisition of competencies related to formal education (e.g., arithmetic, reading, and writing). Domains were administered according to instructions for the BDI-2.

Intellectual Ability

The Verbal Comprehension and Concept Formation subsets of the Woodcock Johnson III (WJ-III) measure crystallised intelligence (G_c) and fluid intelligence (G_f) respectively. These were used to assess further whether performance on FB tests is related to IQ. Tests were administered according to handbook instructions.

Theory of Mind and Subtractive Reasoning

ToM and SR ability were measured by a series of traditional FB tasks comprising six stories presented as cartoon strips (for examples of cartoon strips see Figure 1). Each child was shown the cartoon strips and the stories were narrated by the first author. Two stories involved a person changing location, two involved an object changing location and two did not involve a location change but instead used a false contents test. Tests were presented in that order (for story scripts see Table 2). Earlier work described in the ToM literature has used puppets to convey the story line but a meta-analysis by Wellman, Cross, and Watson (2001, p. 671) found that a story book rather than a puppet presentation of the FB test did not influence children‟s performance. Wellman et al. summarised their conclusion by stating that "knowing that valid and comparable assessments of false-belief performance are uninfluenced by a variety of task specifics, experimenters can confidently vary their tasks over an extended set of possibilities for ease of presentation…" (p. 671). The first four cartoon strips, involving a person or object changing location (much like the traditional Sally-Anne test (Wimmer & Perner, 1983)), were used to test either SR or ToM, dependent on which test question children were asked (see Figure 2).

The final two ToM items involved no change of location. Instead, the false-belief centred on the contents of a box or the nature of an item (see Table 2). These were also presented as picture cards rather than 3D objects. In total, each child was asked four ToM questions and two SR questions. Within the type of item (i.e., person, object, or nothing changing location) test items were counterbalanced for order effects. Children were scored one point for each ToM test question answered correctly. Similarly, children were scored one point for each SR test question answered correctly. Children who answered control questions incorrectly were scored as incorrect for that item. Hence, the maximum ToM score was 4 and the maximum SR score was 2.

Procedure

Most children were tested at preschool or school but a small number (n=8) elected to be tested in their homes, with the parent present. All children were tested in a quiet, distraction-free environment. Testing times varied between mid-late morning and early-mid afternoon. All children completed the full test battery. Testing took about two hours per child, generally over two sessions. The order of test administration was ToM and SR tests, followed by Verbal Comprehension, Concept Formation, and Communication and Cognitive domains of the BDI-2.

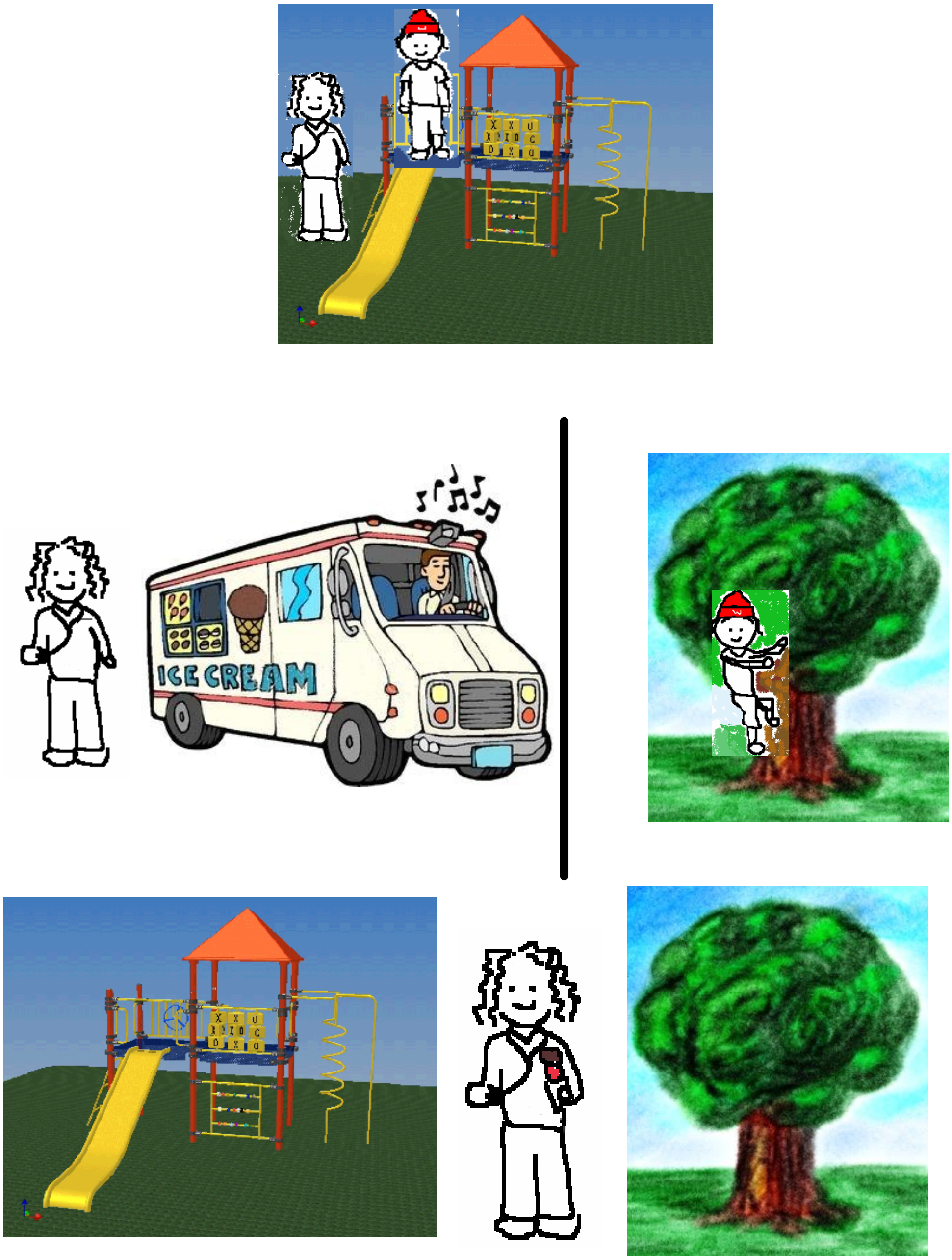

Figure 1. Example of cartoon strip used for both ToM and SR tests. The story is as set out in Table 2, person change, location 2.

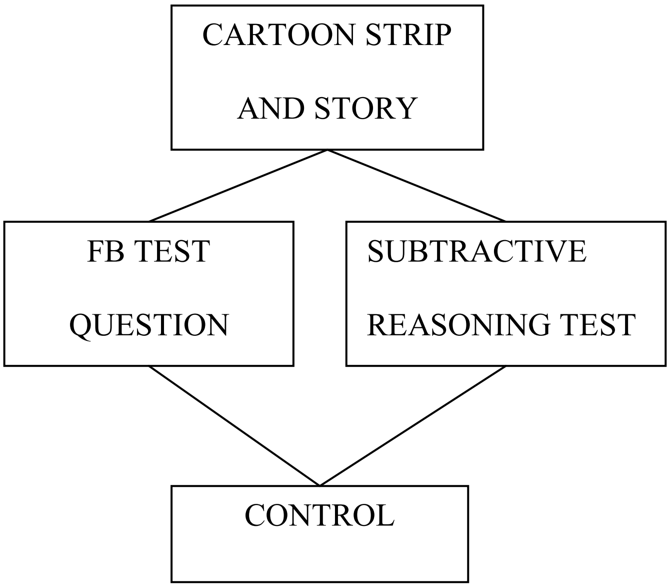

Figure 2. Theory of Mind and subtractive reasoning item presentation.

Table 2

Theory of Mind and Subtractive Reasoning Story Scripts

Person change location 1	Story	Michael and Polly were in the kitchen together. The phone rang and Polly left the kitchen to answer it. While Polly was gone, Michael heard a noise in the bathroom and went to the bathroom to see what it was. Polly has finished on the phone, and wants to speak to Michael.
	ToM test question	Where will she look for Michael?
	SR test question	If there had not been a noise, where would Michael be?
	Control questions	Where is Michael really? Where was Michael before he went to the bathroom? Did Polly see Michael go to bathroom?
Person change location 2	Story	Ben and Sarah were playing together in the playground. Sarah heard the ice cream truck and went to buy an ice cream. While she was gone, Ben saw a cat stuck in a tree and climbed the tree to get the cat down. Sarah bought an ice cream for Ben and one for herself. She wants to give Ben his ice cream.
	ToM test question	Where will she look for Ben?
	SR test question	If there had been no cat, where would Ben be?
	Control questions	Where is Ben really? Where was Ben before he climbed the tree? Did Sarah see Ben climbing the tree?

Object change location 1	Story	This is Lisa and this is Paul. Lisa puts her ball in a basket. Lisa goes to get some lunch. While she is gone, Paul wants to play catch so he takes Lisa‟s ball. When he is finished he puts it in a box. Lisa comes back and wants to get her ball.
	ToM test question	Where will she look for her ball?
	SR test question	If Paul had not played catch, where would the ball be?
	Control questions	Where is the ball really? Where was the ball before Paul played catch? Did Lisa see Paul move the ball?
Object change location 2	Story	This is Mary and this is Tom. Tom puts his chocolate in the cupboard. Tom goes outside to play. While Tom is gone, Mary uses some of the chocolate to bake a cake. When Mary is finished she puts the rest of the chocolate in the fridge. Tom comes back from playing and wants to eat his chocolate.
	ToM test question	Where will he look for the chocolate?
	SR test question	If Mary had not baked a cake, where would the chocolate be?
	Control questions	Where is the chocolate really? Where was the chocolate before Mary used it for cooking?

		Did Tom see Mary move the chocolate?
False contents	Story	Look, this is a smarties container. What do you think is inside the smarties container? Let‟s have a look. There is a pencil inside. (Child‟s friend‟s name) is coming in next and s/he has not seen this smartie container before.
	ToM test question	If I ask (Child‟s friend‟s name) what is inside, what will s/he say is inside the smartie container?
	Control questions	What is really inside the container? What did you say was in the container before you saw that there was a pencil inside? Has (Child‟s friend‟s name) seen that there is a pencil inside?
False object	Story	What do you think this is? Let‟s turn it around and see? It‟s a bird. (Child‟s friend‟s name) is coming in next and s/he hasn‟t seen this picture before.
	ToM test question	If I ask (Child‟s friend‟s name) what this picture is, what will s/he say it is?
	Control questions	What is it really? What did you say it was before you saw that it was a bird? Has (Child‟s friend‟s name) seen that this is really a bird?

Data Preparation

In order to construct language and intellectual ability measures, correlations between BDI-2 and WJ-III raw scores were examined. Correlations are presented in Table 3. When controlling for age, correlations between subtests within the BDI-2 Communication and Cognitive domains were moderate and confirmatory factor analysis verified a single factor for each measure; hence domain raw scores, rather than subtest raw scores have been used in analyses.

Additionally, the relationship between the BDI-2 communication domain and the WJ-III verbal comprehension subset was strong, and Crohnbach‟s alpha for these measures combined was .86, consistent with a single language construct. A composite was therefore derived by standardising the raw scores from each measure to produce a language competence measure.

As seen in Table 3, the correlation between the WJ-III Concept Formation and the BDI-2 Cognitive domain was small and Cronbach‟s alpha for these measures combined was .68, suggesting that they should be regarded as relatively independent. Therefore raw scores for these two measures were used in subsequent analyses.

Table 3

Partial Correlations between WJ-III and BDI-2 Measures, Controlling for Age.

		r	*p*
BDI-2 Communication:	Receptive – Expressive	.48	<.001
BDI-2 Cognitive:			
	Attention and Memory – Reasoning and Academic	.46	<.001
	Attention and Memory – Perception and Concepts	.47	<.001
	Reasoning and Academic – Perception and Concepts	.58	<.001
Language competence:			
	BDI-2 Communication – WJ-III Verbal Comprehension	.71	<.001
Intellectual Ability:			
	BDI-2 Cognitive – WJ-III Concept Formation	.38	.003

Note: N = 60

Results

Sample Descriptive

Language and Cognitive Development and Intellectual Ability

The standardised scores, for the children who participated in this study, were slightly above average but normally distributed for WJ-III and BDI-2 measures. Mean, SD, and range for each measure are presented in Table 4.

Table 4

Standardised Scores for the WJ-III and BDI-2 Measures.

	M	SD	Min score	Max score
WJ-III Verbal Comprehension Subset	112.02	14.24	86	160
WJ-III Concept Formation Subset	106.63	14.11	74	135
BDI-2 Communication Domain	115.07	11.22	82	145
BDI-2 Cognitive Domain	110.43	10.52	80	132

Note: N = 60

Theory of Mind and Subtractive Reasoning

Table 5 displays M and SD for each of the FB task types (unexpected object/contents, object change of location, person change of location) and the SR tasks (object change of location and person change of location). To examine whether performance across the tasks was consistent correlations for test items were generated. Given the nature of the variables Phi correlation statistic or Cramer"s V were used where appropriate. Only correlations within the two types of tasks were significant; that is, between the two change of location tasks (ϕ = .39, $p < .01$) and between the two unexpected identity tasks (ϕ = .47, $p < .001$). This suggests either that the two types of items, false objects/contents and change of location, were not testing the same construct; or more likely that these types differed in degree of difficulty.

Table 5

Mean Score for Each FB and SR Task.

		M	SD
FB	ToM person location change	.12	.32
	ToM object location change	.35	.48
	ToM unexpected contents	.73	.45
	ToM unexpected object	.77	.43
SR	Person location change	.40	.49
	Object location change	.53	.50

Note: N = 60; possible score for each ToM task 0,1

To examine whether ToM test performance was positively related to language competence, intellectual ability, and subtractive reasoning ability, correlations between ToM and the language competence measure, WJ-III Concept Formation and BDI-2 Cognitive domain, and SR measures were examined. As seen in Table 6, ToM performance was correlated with language and subtractive reasoning, after controlling for age. This lends support to the hypothesis that ToM is related to language and reasoning. However, no support was found for a relationship between ToM and overall intellectual ability.

Table 6

Partial Correlations between ToM and Language, Intellectual Ability, and Reasoning Measures Collapsed for Age Group, Controlling for Age.

	r	*p*
Language competence	.36	.005
WJ-III Concept Formation Subset	.15	.26
BDI-2 Cognitive Domain	.18	.17
Subtractive Reasoning	.36	.005

Note: N = 60

To examine whether correlates of ToM performance were consistent across age groups, further analyses were carried out on the 4-year-old and 6-year-old samples separately. As can be seen in Table 7, in 4-year-old children ToM tended to improve with age within the cohort, and correlated significantly with language and G_f (as measured by WJ-III Concept Formation subset). These results therefore indicate that within this age group ToM performance is potentially sensitive to individual differences in these measures.

However, as can be seen in Table 7, in 6-year-olds ToM was only correlated with SR. Insofar as correlations, within the cohort, between ToM, age, the language and intellectual ability measures were very small and were not statistically significant, it appears that by 6 years of age individual differences in language may be less relevant to the ongoing development of ToM than individual differences in subtractive reasoning skills.

Linear regression analysis using the enter method examined the relative extent to which language, intellectual ability, and SR measures accounted for ToM performance across the whole sample. Variables were entered in order from most to least correlated (i.e., language competence, SR, BDI-2 Cognitive, & Concept Formation). Only language competence (B = .034, β = .47, p <.001) and SR (B = .43, β = .29, p =.012) contributed significantly to ToM. Together, these explained 42% of variance in ToM [R^2=.42, F(2,57) = 20.96, p <.001]. Additionally, separate regression analyses were performed on each age group to determine whether predictors of ToM change with age. In 4-year-olds only language competence (B = .05, β = .54, p =.002) predicted ToM performance [R^2 = .30, F(1,28) = 11.73, p =.002]. However, in 6-year-olds, only SR (B = .50, β = .38, p =.04) predicted ToM [R^2 = .14, F(1,28) = 4.65, p =.04].

Table 7

Correlations between ToM and Age, Language, Intellectual Ability, and Reasoning Measures.

	Correlation with ToM			
	4 years (n = 30)		6 years (n = 30)	
	r	*p*	r	*p*
Age in months	.37	.045	.055	.77
Language competence	.54	.002	.14	.47
WJ-III Concept Formation Subset	.36	.048	-.07	.71
BDI-2 Cognitive Domain	.33	.076	.098	.61
Subtractive Reasoning	.18	.33	.38	.04

Finally, to confirm a difference in the relationship between cognitive abilities and ToM across the age groups, principal components factor analyses, using promax rotation, were performed on the 4 and 6 year old data separately. Two factors, cumulatively explaining 73% of variance, on which ToM loaded almost equally, were found in 4-year-olds. The first factor was well described by the composite language measure and the second by counterfactual reasoning. By 6 years of age these two factors explained 64% of variance and ToM was now only explained by the factor encompassing SR. The first factor, accounting for 38% of variance is well described by the language and intellectual ability measures; the second, accounting for a further 27% of variance, captures both the necessary precondition of SR and ToM. Table 8 displays these results. Thus, there were differences in the relationships between ToM and language, intellectual abilities, and SR for the different age groups.

Table 8

Principal Component Analysis for ToM, SR and Language and Cognitive Measures in 4-year-olds and 6-year-olds

	Factor Loadings			
	4 years		6 years	
	Developmental factor	Counterfactual reasoning factor	Developmental factor	Counterfactual reasoning factor
ToM	.60	.64	.07	.79
SR	.14	.90	.11	.86
Language Competence	.91	.40	.81	-.029
WJ-III Concept formation	.81	.31	.77	.16
BDI-2 Cognitive domain	.81	-.01	.76	.12

Discussion

As expected, we found that ToM correlated positively with language competence and subtractive reasoning measures when considering 4-year-old and 6-year-olds combined (Milligan, et al., 2007; 2000; Slaughter & Repacholi, 2003). When considering the age groups separately, however, the relationship between ToM and development was more complex. In the 4-year-old age group language was related to ToM, which is in keeping with what is already known of the relationship between language and ToM (Astington & Jenkins, 1999; Happé, 1995). Furthermore, a distinct factor, proposed to be a developmental component, underlying performance across language, intellectual ability and ToM, was detected in the 4-year-old group, along with a separate factor, proposed to be a counterfactual reasoning component, equally explaining ToM, but mostly consisting of SR. This fits with theories that propose a single factor underlies performance on intellectual tasks in young children (Anderson, 1992). In short, ToM performance in 4-year-olds was dependent on some general developmental factor, largely resembling language acquisition, and also on the SR precondition involved in ToM reasoning. The consistent finding that ToM is related to language can thus be largely explained by intelligence theory (Anderson, 1992). Indeed, Shanker‟s (2004) proposal that children‟s early learning about the social world is underpinned by the sharing of experiences with a caregiver, applies equally to children‟s early learning about the physical world. A module specifically designed to process social information may not be necessary in a typically developing child.

In 6-year-olds the relationship between ToM and language, intellectual ability, and SR measures was more complex. In this group only SR remained related to ToM performance. Factor analysis in the 6-year-old group showed that two distinct factors contributed to performance across the tests. One factor was well described by language and intellectual ability measures in the battery; and the other underpinned ToM and SR. This is a new finding, primarily because this is the first research that has examined cognitive skills in relation to ToM in older children. The most closely related published research has investigated the process underlying belief-desire reasoning in children and adults (Friedman & Leslie, 2004; Leslie, et al., 2005). Friedman and Leslie (2004) proposed that there is a developmental shift in the processes underlying reasoning about beliefs and desires. The current finding, that ToM in 6-year-olds does not look much like ToM in 4-year-olds, is plausible in light of this proposed developmental shift; although Friedman and Leslie (2004) suggest that this shift happens much later in childhood. The current findings also lend some support to findings that ToM correlates with IQ in children with autism (Muris, Steerneman, & Merckelbach, 1998). When ToM does develop in children with autism it is likely that intellectual and reasoning abilities facilitate ToM task performance. It is likely too that delays and differences in language development in children with autism will distort the earlier language-based success in FB tasks seen in 4-year-olds here.

In this study the distinction between ToM competence and FB task performance was apparent. Children appear to have found FB tasks involving a character changing location more difficult than those involving an object changing location, and those FB tasks involving no change of location the least difficult. Those children passing the unexpected object/contents FB tests but not passing the change of location FB tests displayed ToM competence (understanding that other people do not know everything they know) but demonstrated task performance limitations when answering change of location FB test questions.

In fact, the FB test battery used in this study appears to have been unusually difficult for children, thereby, inadvertently emphasising the performance component of ToM tasks. Indeed the literature suggests that by age 4 half of all children perform above chance on ToM tests which involve an object changing location, and that by age 7 or 8 ToM tests show ceiling effects (Slaughter & Repacholi, 2003). However, in the current study only four out of 30 four-year-olds and approximately half of the 6 year olds performed above chance on the ToM test battery. This was mostly due to children performing very poorly on the FB test items involving people changing location. Similar stories have been used in previous research without increasing task difficulty (2000; Riggs, Peterson, Robinson, & Mitchell, 1998). It must be noted that these were not more complex FB tasks and that the only difference was that, rather than an object moving while a character was absent (as is the case in the Sally-Anne FB test), a person moved without the character knowing. It is possible, therefore that, despite Wellman et al.‟s (2001) conclusion that small task variations in test question or story presentation do not increase task difficulty, the cartoon strip presentation rather than a puppet presentation did increase difficulty of the tasks for even 6-year-old children in this study. Further research, perhaps using repeated measures of these types of variations, story book versus puppets, in tasks using children across a broad age range, may be able to determine whether the presentation method of FB stories contributes to FB test performance in children, and the extent to which this is related to age. In any case, if FB tasks are to be used in future research, a standardised task ought to be developed.

In conclusion, when examining what area of FB tests children have difficulty with, it is important to distinguish competence from performance (Wellman, et al., 2001). Certainly, the development of ToM competence relies on language-based interactions between child and caregiver and performance on FB tests relies on language comprehension (Astington & Jenkins, 1999). However, current results suggest that, as children‟s cognitive abilities mature, successful reasoning about people‟s thoughts and beliefs becomes less reliant on language ability. In populations that lack the necessary language and social interactions required to develop ToM competence, perhaps developing reasoning strategies, which this study suggests are used by older children on more difficult tasks, can help deliver ToM competence. Whether ToM strategies will alleviate social difficulties is an altogether different discussion (Astington, 2003).

CHAPTER 5 - Reliability and Validity of Theory of Mind Measures

The ToM battery used in the study, presented in the previous chapter, was found to be unusually difficult. This unexpected and perplexing difficulty was unlikely to be the result of the sample alone, because the children in the sample showed above average development and IQ across all domains tested. Hence, something about the test battery items or the ToM test procedures is likely to have caused the test battery difficulty. For this reason, it was deemed appropriate to examine in more detail the ToM literature concerning reliability and validity, before presenting a study that explored children"s difficulty with the test battery described in Chapter 4. In short, this chapter reviews reliability and validity of Theory of Mind measures as reported in the literature.

While the literature discussed in this chapter varies in regard to the way in which reliability and validity are defined, the current author"s definition of these terms is provided here by way of emphasising the aims of the current chapter. Test-retest reliability is defined as the propensity of a measure to provide similar scores for an individual across multiple times. Validity is discussed as the ability of an experimental measure to accurately capture and evaluate the intended construct/s assessed by the measure.

Little research addresses test-retest reliability and convergent validity of commonly used tests of ToM, despite such tests being in use for more than two decades. Reliability and validity of ToM tests are likely to be related to a number of factors; three key factors are discussed here. Firstly, reliability is likely to be better when more items contribute to a ToM score. Nevertheless, ToM research often employs a single item measure of ToM or a small number of false-belief (FB) type items to assess ToM. Secondly, task reliability can be affected by variations in administration of tests designed to measure ToM. Procedures for testing ToM vary in a number of potentially important ways (Wellman, Cross, & Watson, 2001). Thirdly, reliability can also be related to general performance variability of young children and children from clinical populations, who make up the majority of the ToM research population.

Recent research summarised by Hutchins, Prelock, and Chace (2008) called for ToM test batteries, that are sensitive to a range of aspects of children‟s understanding of the mental world, to control for effects of poor task or performance reliability. An example of such a battery was developed by Wellman and Liu (2004). These authors scaled seven items in a theory of mind test battery spanning a number of aspects of mental state understanding. Their scale included tasks designed to measure understanding of: diverse belief, false belief (both unexpected contents and change of location tasks), diverse desire (the understanding that different people can have differing desires), belief emotion (how someone will feel given a mistaken belief), knowledge, and real-apparent emotion (person feels one thing but displays another). Despite being an advance on single item assessments of ToM, the battery still only measured ToM with a small number of items. Moreover, the oldest and most frequently used measure, the FB test, which has been found to vary greatly in reliability and validity, is still commonly used.

In addition to issues of test-retest reliability and convergent validity, the ToM literature also includes debate about the construct validity of ToM tasks. *Theory and Psychology* (2004) devoted an issue of their publication to critical assessment of Theory of Mind. This journal issue contains the largest published body of work that takes issue with the typical form of measuring ToM, namely the FB task, and the underlying assumptions of the populous of ToM research. An introduction to the issue states that

> "The ToM framework has been associated with the fastest growing body of empirical research...yet the approach has so far escaped a serious assessment or historical investigation, and has avoided any sustained engagement with alternative explanations of social action, such as discursive psychology, ethnomethodology or sociocultural psychology. By comparison with the wealth of articles and books published by proponents of ToM, the critical

> literature on ToM is rather small, consisting of a few articles and monographs taking issue with assumptions, methods and findings. The limited published criticism has been simply ignored." (Leudar, Costall, & Francis, 2004, p. 572)

While Leudar et al. take a strong stand against the lack of engagement with alternative perspectives; this researcher has not found a complete disengagement in the literature with criticisms of ToM. The ToM literature has attempted to consider the tests used in ToM research that may help to make sense of the varying findings related to ToM competence in typically developing children. Mostly, however, these methodological aspects are in relation to test-retest reliability and consistency, rather than construct validity. The current chapter reviews these research findings as well as the main criticisms made of ToM accounts.

Test-Retest Reliability

The six published studies examining test-retest reliability of ToM measures report mixed findings about the reliability of ToM tasks. These studies differed in methodology, types of ToM tasks assessed, and populations sampled. The following is an overview of the ways in which these studies differed and how these differences might account for reported differences in reliability. The section is divided by non-clinical and clinical samples for ease of collating the range of information.

Non-Clinical Samples

The first test-retest reliability study, conducted by Mayes, Klin, Tercyak, Cicchetti, and Cohen (1996), reported unacceptably poor test-retest reliability. While the study used a sample of children which reflected the most common cohort of interest in ToM research (children 3 to 6 years); the study used a number of non-standard elements to assess ToM. Firstly, false-belief tasks were presented in non-standard ways, using video footage of real people in place of props or story book formats. Secondly, the study analysed data for both traditional ToM test questions (i.e., predicting action of character) and non-standard questions (e.g., asking what the character thinks).

In order to assess whether reliability differs for different age groups, the sample of children was divided into two age groups; the first comprised 11 children aged 4 years (M=42.8 months, SD=3.9) and the second was made up of 12 children older than 4 years (M=55.8 months, SD=6.6). Mayes et al. hypothesised that after the age of 4 years children are more likely to succeed on FB tasks and that before this time children are more likely to be variable in their performance. Counter to expectations both age groups displayed unreliable performance for the three ToM tasks across a test-retest period of 2 - 3 weeks. Moreover, poor reliability was found for both the test questions (aimed to assess ToM competence) and also for the control question (designed to measure children‟s story comprehension). Detailed examination of factors that were likely to contribute to the poor reliability reported by Mayes et al. follows.

Mayes et al. used a number of tasks that differed in presentation format. Importantly, reliability varied across tasks that differed in presentation format. Therefore, it is likely that differences in task structure affected task reliability. The most reliable task was the traditional Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985) where the story was simultaneously narrated and acted out. The two remaining tasks did not utilise narration and, instead, actors spoke. For instance one actor tells the other that she has a new teddy bear and when this same actor returns and wishes to play with the bear, which unbeknownst to her has since been moved, she exclaims that she now wishes to play. Arguably, following the story in non-narrated events involves cognitive processes that differ from those involved in viewing narrated events. In narrated events children simultaneously listen to the story while watching it unfold. The narrator explicitly draws the child‟s attention to pertinent information rather than allowing the child to interpret relevance of events. For instance the narrator tells the child "This is Sally and this is Anne. Sally has a teddy bear and puts it in the green box."

Examining the distribution of scores across the three tasks, the non-narrated tasks appear to have been marginally easier than the narrated tasks for younger group of children. Worryingly, consistent performance on the traditional narrated task may have been an artefact of task difficulty rather than task reliability. That is, the task was consistently too difficult for this young age group with 10 out of the 11 children failing it on both testing occasions. In contrast, the corresponding control question, which assessed the child‟s understanding of Sally‟s knowledge in relation to the events she did not witness, was not found to be reliable and demonstrated much more variable performance; only one child failed the question at both times, five children passed at both times, and 5 children passed at one of the two testing times. The same can be said for the other unreliable items. Generally unreliable items showed greater variance in group performance; suggesting that, in this age group, the tasks were largely cognitively inaccessible.

Alternative explanations can also account for poor reliability. For example, poor reliability might be due to maturation or practice effects, because there was a relatively short time between test and retest (2-3 weeks). Maturation or practice effects cannot, however, have been entirely responsible for poor test-retest reliability, given that performance on the second test was worse for some children. A more likely explanation is that children in this age group are cognitively limited and are, therefore, prone to unreliable performance. Mayes et al. suggested that poor test-retest reliability may have been explained by difference between children‟s verbal abilities, a variable that was not controlled for in their study.

A later study, using twice as many children, attempted to control for the non-standard elements of the Mayes et al. study (Hughes, et al., 2000). Hughes et al.‟s methods resulted in findings of better reliability. Hughes and colleagues proposed that Mayes et al.‟s nonstandard presentation format and unusual scoring methods (including scores of children who passed FB questions but failed control questions) may have contributed to findings of poor test-retest reliability. Hughes et al. also proposed that nonstandard presentation, used by Mayes et al. may have added nonspecific task demands.

There were three important differences between the Mayes et al. and Hughes et al. studies that might have contributed to differences in reported reliability. Firstly, Hughes et al. used a longer test interval of four weeks, and their sample was approximately 10 months older. The longer test interval might have reduced practice effects. Moreover, the older sample may have improved reliability by decreasing inconsistencies due to cognitive immaturity. The second difference was that Hughes et al. scored children who failed control questions as failing the test questions, while Mayes et al. measured reliability across both test and control questions. Grouping children who failed control questions with those who failed both control and test questions controls for inconsistent performers, thereby improving reliability. The third factor that might account for Hughes et al.‟s better reliability was that these authors used standard narrated stories. Standard narrated stories were found to have the best reliability in the Mayes et al. study.

In addition to using methods that might improve reliability, Hughes et al. also attempted to control for differences in reliability in relation to children‟s intellectual ability. The authors reported that there was no difference in ToM task reliability in relation to children‟s intellectual ability. In part this might be related to the intellectual ability measure used. Hughes et al. controlled for IQ, which has not been reliably linked to ToM performance. Given the reliable relationship between language and ToM in the age group used in their study (as shown by Happé, 1995), a more suitable comparison would have been for verbal ability rather than IQ, as suggested by Mayes et al.

The Mayes et al. and Hughes et al. studies also differed in a final and important way. The studies conclude reliability using different statistical parameters. Hughes et al. reported only Kappa values, unlike Mayes et al. who reported more stringent PO, P_{neg} and P_{pos}; statistics which are more suitable for data where occurrence of one event (passing or failing) is very high or very low (citing Cicchetti & Sparrow, 1981). According to these criteria a task can only be considered reliable when Kappa is greater than .40, and PO, P_{neg} and P_{pos} are each greater than .70. Kappa across the tasks used by Hughes et al. ranged from .29 through to .72 with most values falling around .50.

Clinical Samples

Three further studies examined reliability in clinical samples, again producing mixed findings. These studies are compared here. The first study of the reliability of ToM tasks in a clinical population was conducted by Charman and Campell (1997). The authors examined internal consistency or convergent validity, that is whether different tasks measure the same construct, in a single session for children with a learning disability. Using the more stringent measures of reliability employed by Mayes et al., they too found unacceptable reliability for some FB tasks and belief-desire reasoning tasks. The tasks that were judged as reliable were moderately reliable or fell short of meeting stringent reliability criteria by .01 Kappa.

In contrast to earlier findings, that reliability was not related to cognitive ability, (Hughes, et al., 2000), Charman and Campbell found better reliability in relation to increased verbal ability. The sample consisted of 36 children and young adults with Down Syndrome or a learning disability of unknown organic aetiology (age range 7 years 7 months – 19 years 6 months). Despite the chronological age of the sample the mean verbal mental age (VMA) (M = 4 years 11 months) and non-verbal mental age (NVMA) (M = 5 years 1 month) were relatively low. Participants completed three standard FB tasks and two belief-desire tasks, all assumed to measure ToM. In contrast to Hughes et al., Charman and Campbell found that children who did not perform reliably had statistically lower VMA and NVMA than children who reliably passed the ToM tests. This may be an artefact of the clinical population tested, in that children who test as less developmentally mature are perhaps generally less consistent in their performance across a wide range of tasks compared to children who test as developmentally more mature. The difference does, however, support Mayes et al.'s supposition that children who perform unreliably do so because of task specific demands, rather than a lack of ToM. Furthermore, if ToM is to be considered a developing skill, rather than an all-or-nothing skill, some variation in performance while acquiring this skill is to be expected. Interplay of processing demands, specific to ToM tasks but not related to ToM, and variation in ToM competence are both likely to contribute to unreliable performance in pre-school aged children and clinical populations.

Further support for the proposition, that reliability of ToM tasks is related to non-ToM aspects involved in completing ToM tasks, comes from two studies that have examined performance stability in samples of children with autism spectrum disorders (ASD). The first of these studies found good reliability in a sample of verbally mature children (Grant, Grayson, & Boucher, 2001) and the second found that reliability varied in relation to verbal maturity (Hutchins, et al., 2008).

Grant et al. (2001) used the same research methodology as Charman and Campbell (1997). By testing 22 children on a number of tasks in a single session, Grant et al. assessed convergent validity of tasks rather than test-retest reliability. Seventeen children performed reliably, with 12 consistently passing and five children consistently failing. Performance was compared across four standard FB tasks, generating six task comparisons; the Sally-Anne task, two deceptive box tasks (unexpected contents), and a change of location task which used three boxes as possible locations for the hidden object rather than two locations as for the Sally-Anne task. Using the stringent reliability measures of Kappa PO, P_{neg} and P_{pos}, all but one task comparison (the Sally-Anne task vs. the first deceptive box task) showed good reliability (Kappa range = .59 - .93). A likely reason for differences in convergent validity reported by Grant et al. and Charman and Campbell is that Grant et al.‟s sample had a much higher VMA (M = 8 years 5 months). That is to say, consistent responding for a specific task is not just related to the reliability of the task, instead consistent responding can also be depend on the reliability of children in general, where children‟s reliability is associated with their overall maturity (in this case verbal maturity). Consequently, these findings suggest that these tasks measured the same construct reliably when used with verbally mature children with ASD.

Less reliable test-retest performance has been evident for children with ASD who have low levels of verbal maturity (Hutchins et al., 2008). Hutchins et al. tested 17 children (mean

VMA = 6.7 years; range < 2 years – 15.25 years) across both short (2-7 weeks) and long (8-16 weeks) intervals. Much like the methods used by Mayes et al., Hutchins et al. used multiple questions (total of 16 test questions) to target ToM ability over nine tasks. Over a test interval ranging from 2 - 16 weeks, the authors reported moderate to good agreement for 10 of the 16 questions, unacceptable reliability for three of the remaining questions, and good agreement for three more questions only if either P_{neg} or P_{pos} were disregarded. Hutchins et al. stated that disregarding either of these was acceptable if a task was found to be very difficult or very easy. The study attempted to examine differences in verbal ability between children labelled consistent or inconsistent performers, but found no significant difference. Without further information it is difficult to assess whether this was an artefact of the small sample size for the study. An absence of details outlining how many children were consistent or inconsistent performers precludes an assessment of the power of the analyses employed to detect differences in verbal mental age, should such differences exist. Nonetheless, the less verbally mature sample of children tested by Hutchins et al. might account for some of the differences in reliability across the two studies examining reliability of tasks for children with ASD. Indeed, Hutchins et al. appeared to conclude that ToM test-retest reliability was related in some way to children‟s non-ToM abilities. They stated that, for the most verbally limited children in their study, "language ability, pragmatic understanding of the testing situation, motivation or other constructs irrelevant to ToM competence" (p. 204) were more likely to influence performance on the ToM tasks. Although Hutchins et al. only conceded the relationship between maturity and test-retest reliability for the most verbally limited children in their sample, extending this proposed relationship to children who have somewhat less but still limited language ability, memory and attention capacity is not unreasonable, and studies with sufficiently large samples to test this relationship support this supposition (Charman & Campbell, 1997).

To summarise the findings in relation to test-retest reliability and convergent validity of ToM measures, limited studies, using varying methods, sampling from vastly different populations and often using small samples, have reported varying degrees of reliability. Furthermore, although studies have not sufficiently determined the contribution of language ability to performance reliability amongst young children, it seems that children who perform inconsistently are likely to have poorer language skills and perhaps also lower overall cognitive functioning. Non-trivial processing demands created by artefacts of task administration are also likely to contribute adversely to reliability of ToM measures.

Consistency

Non-Clinical Samples

The literature reports a large range of ToM performance for young children. This has resulted in debate about the consistency of the standard measures used in research. Consistency of the standard measures has been discussed in relation to how these measures have been adapted and changed across studies. The most extensive assessment of consistency for ToM tasks in typically developing children comes from a meta-analysis by Wellman, Cross, and Watson (2001). Their analysis of data from 143 studies, including 591 ToM task conditions, set out to account for inconsistencies in measured age of ToM onset across studies and to answer questions raised in the literature about what ToM tasks measure. Main questions were, whether ToM tasks mask early ToM competence, due to undue difficulty, or whether ToM undergoes a distinct conceptual change in early childhood. Wellman et al. summarised the variations made to ToM tasks in the literature that attempted to reduce task difficulty in order to expose earlier ToM. Wellman et al. argued that in order to distinguish between the two main claims, masking competence versus distinct conceptual change, performance in relation to task manipulations must evidence a main effect (e.g. an increase in performance regardless of age) and an interaction with age (e.g. the effect disproportionately benefits younger children). Furthermore, Wellman et al. stated that the task manipulation must result in younger children performing above chance.

While findings in line with Wellman et al."s proposed analyses might distinguish between alternate accounts of ToM, interaction effects (between a main effect and age) are not necessary to conclude that ToM ability is masked in younger children. The reasoning behind Wellman et al."s proposition is that task manipulations should account for performance of younger children relative to older children; therefore, according to Wellman and colleagues a task manipulation that equally benefits both older and younger children only shows that the task was more difficult than it ought to have been. The problem with this argument is that it relies on an all or none account of ToM, where it is assumed that once a child has acquired ToM s/he will reliably demonstrate this capacity regardless of task complexity. However, when performance on a task is improved for children, irrespective of age, in response to a particular manipulation, it shows that the task is unduly complex for all ages, and this does not preclude the conclusion that it masked young children"s ToM. Unless a task shows ceiling effects for older children, then all children regardless of ability level ought to perform more reliably on less complex tasks than on more complex tasks. Arguably for early competence accounts to be substantiated using empirical methods all that is required is that very young children reliably demonstrate ToM on an age appropriate task.

Non-empirical methods demonstrating ToM in younger children have not been addressed in the Wellman et al. meta-analysis, but these should not be discounted. Such methods might include observation of naturalistic behaviour and language that reflects an understanding that other people have minds distinct from one"s own mind. A more detailed discussion of these possibilities will appear toward the end of this chapter.

Leaving behind for a moment the reasoning behind their analyses, Wellman et al."s findings do tend to support the proposition that younger children do not possess a working ToM. To aid analysis of their data set Wellman et al. transformed original proportion correct data using logit (ln natural logarithm) in order to: "[yield] a linear relation that allows systematic examination of the data via linear regression; ...[eliminate] restricted range inherent to proportion data; and... [yield] a dependent variable and measure of effect size that is easily interpretable in terms of odds and odds ratios" (p.662). The transformed data clustered broadly along a proposed developmental trajectory where probability of correct responding increased with age. Despite finding a number of task variations that improve ToM task performance (motive of character in the story, participation of child in the task, real presence – whether the target object was present and real or had been consumed, salience of the protagonists mental state, and country in which the child resided), the analyses found only one ToM test variation that interacted with age. This was for ToM tasks that included a temporal marker in the test question (e.g. "When Sally comes back, where will she look first for her ball?"). The task manipulation was found to have benefited older children but not younger children. Wellman et al. explained this finding in relation to the complexity of the test question. They supposed that adding the temporal marker „first" increases sentence complexity for younger children but that older children have the cognitive resources to use the extra information when answering the question. However, a limitation of the meta-analysis is the age range of children in the included research. Certainly this limitation stemmed from a lack of research conducted on children outside the period of developmental interest, that is, the age at which ToM begins to emerge. The studies included mostly examined children‟s abilities between the approximate ages of 3.5 to 5.7 years. Only a few studies either side of this age range were represented and these did not necessarily align neatly with the body of research.

Another limitation of Wellman et al.‟s analysis is that outliers were not all identified and discussed. Nevertheless, the outlier most troubling for Wellman‟s theory that children undergo a distinct conceptual change was explained. The outlier was from a study that found that children aged 30 and 36 months passed a ToM task. According to Wellman et al., success on the task simply involved the child repeating what s/he was told. More specifically, Wellman et al. reported that the ToM task involved the character saying "I think my cookie is in the red box" after which, according to Wellman et al., the child is immediately asked "Where does (character) think his cookie is?" Wellman et al. suggested that all the children needed to do was to repeat what they last heard and as such this measure was likely to artificially inflate ToM success. Unfortunately, the study was presented at a conference and not published. For that reason an independent assessment in this chapter of that study‟s materials and procedures is not possible. While the task may have been very easy, it is likely that it was able to discriminate to some degree between 2-year-olds; for whom it might not have been so straight forward. For instance, despite the apparent ease of the task, a 100% success rate was not evident from the data presented in the meta-analysis. The proportion correct appears to have been around .88. Nevertheless, it is likely that the task was excessively simplified and was not a true measure of ToM.

More troubling were unidentified outliers that show children between approximately 3.3 years and 4.2 years consistently failing tasks, children aged just over 6.7 years performing at chance, and children above 4 years performing below chance. While data in these studies might not have been able to be easily explained, at the very least the studies from which these data came should have been identified. The pre logit transformed data showed probability of 4-year-old success ranging from zero to 0.9. The logit transformation reduced the spread of the data but still left a wide band around the developmental trajectory for all ages. A meta-analysis of this sort can statistically show that, on average, performance over a large number of studies will find an age trend, but variation in ToM performance across tasks will still be problematic for isolated smaller studies.

A further limiting factor of the meta-analysis was that in order to account for excessively difficult tasks or “extremely confused children” the meta-analysis excluded from their sample studies for which fewer than 60% of children passed control questions or for which 40% or more children originally tested were excluded (p. 659). Although this measure was justified as quality control and only nine out of 178 potential studies were eliminated for this reason, the restriction may have eliminated cases with contrasting findings. An analysis, or at the very least a summary, of the findings of the eliminated studies ought to have been reported, to justify their exclusion.

To summarise, Wellman et al. showed that on average there is a developmental trend related to ToM measures, regardless of task variations. Findings of the research indicate that inconsistencies of performance are problematic for ToM research. Wellman et al. argue that these differences do not pose a problem for the theory behind ToM tasks, but these differences do pose a problem for the interpretation of research that utilises tasks without having a standardised assessment of the task they employ.

Clinical Samples

To date there has been no systematic investigation of the consistency of ToM tasks for children with autism. The most relevant findings that address, to some extent, the issue of consistency for children with autism, come from a relatively large study, in contrast to the small sample sizes of most studies, that sought to account for differences in ToM performance within a sample of children with autism (Happé, 1995). Happé"s (1995) study included pooled data from a number of studies originating from a single research centre collected over five years; details of which were reported in Chapter 2 of the current thesis. The main aim of the study was to account for individual differences in children"s ToM performance, in terms of verbal maturity and verbal IQ. Unlike Wellman et al., Happé only included data from the standard Sally-Anne task and an unexpected contents (Smarties) task. At the time of Happé"s publication, 27 previous studies had examined ToM in children with autism and, as summarised by Happé, success across the 16 studies that used standard ToM tasks ranged from 15 – 60 % of children with autism passing tasks. Sample sizes for the studies were typically small (most including fewer than 20 children). Therefore, in her study, Happé sought to overcome the difficulties related to small sample sizes. Children"s verbal ability was measured using the British Picture Vocabulary Scale (BPVS), generating both a verbal IQ (VIQ) score and a verbal mental age (VMA) score. VMA reflects the approximate age at which the child performs on the test and VIQ reflects how well the child performs in relation to his/her age matched peers. As discussed in Chapter 2, Happé reported mixed findings relating to verbal IQ (VIQ), verbal mental age (VMA), and chronological age for three samples of children; normally developing children, children with autism, and children with an intellectual disability. Specifically, the results were not very clear for typically developing children in the study. The main finding of the study (as has been discussed in Chapter 2) was that children with autism required, on average a higher VMA to pass ToM tasks than did children without autism.

Consistency of performance across the two ToM measures used in the study varied between the three groups. Approximately even proportions of children with autism (18/70) and children with an intellectual disability (9/34) performed inconsistently across the two ToM tasks; passing one task but not both. Comparatively fewer typically developing children (11/70) performed inconsistently on the ToM tasks. These findings of variation in performance consistency fit with research related to test-retest reliability presented earlier in this chapter; whereby, children who are more cognitively competent tend to perform more consistently. Happé concluded that inconsistent ToM performance appeared to be related to task difficulty. She argued that differences in task difficulty were not a product of the ToM element of the task, and that in this way neither test was "process pure" (p.851). Furthermore, Happé elaborated that performance consistency appeared to be related to VMA. Happé compared VMA scores of children who consistently passed or failed both ToM tasks to those who passed one task but not both, in order to examine the potential role of VMA in passing ToM tasks. From this comparison, Happé concluded that children with autism who performed inconsistently appeared to do so because of chance, given that on average the VMA of this group was most similar to those children who consistently failed both tasks. Conversely, Happé concluded that children in the other two groups who passed only one task appeared to be reflecting a difference in difficulty between the tasks. That is, that these children did have ToM but that non-ToM-related task difficulties hampered performance, given that these children were similar in VMA to consistent passers. Although this conclusion is logically formed it is likely that, in all three groups, children who performed inconsistently could have either experienced task difficulty for the task they failed or guessed correctly at the task they passed. Curiously, children appeared to find the Smarties task most difficult. This task is much less language rich than the Sally-Anne task, given that it does not involve a story in which a number of events take place. Happé did not outline the method of presentation of these tasks and it is therefore difficult to resolve why this difference in difficulty was found. Furthermore, and perhaps most importantly, no statistical analysis of the difference in apparent task difficulty was presented and it is, therefore, not possible to gauge whether this was a real difference.

The key finding of this study of a relationship between VMA and ToM was the first definitive finding in this field of literature. The study‟s findings include a number of peculiarities that are difficult to explain, other than to invoke non-specified sampling artefacts. More relevant to this discussion is, however, that the study‟s findings can be used, to some extent to address the consistency of ToM measures when used with children with autism. To this end the findings suggest that ToM performance across differing tasks is less consistent in children with autism when compared with typically developing children.

In total, the Wellman et al. and Happé studies indicate that more research is needed to clarify the factors affecting consistency of performance for ToM measures. Most important is an examination of the varying findings for ToM in populations with autism. Although Happé‟s findings suggest that verbal ability is a key factor, the study did not measure task artefacts. Research measuring the impact of common task variations, found in the literature, using a sample of children with smaller range of verbal ability, would be useful in examining consistency of ToM tasks in children with autism.

Theory of Mind Test Batteries

Non-Clinical Samples

Concerns about the reliability, consistency, and construct validity of single or limited multiple task measures of ToM have resulted in a number of attempts at creating a multiple item ToM test. Results of these studies are presented here.

Muris et al. (1999) examined the validity of a ToM test developed by Steerneman (1994, as cited in Muris et al.)[1] which was designed to test the five stages of ToM development proposed by Flavell, Miller and Miller (1993). These five stages have been described by Muris et al. as being embedded in three main phases: "precursors to theory of mind; first manifestations of a real theory of mind; and mature aspects of theory of mind" (p. 69), with three of the five stages comprising the first phase, precursors to ToM. The first stage is the understanding of mind, that is, children develop an appreciation of the needs, emotions, desires, and so forth, of other people. Secondly, children acquire an understanding of the connection between the mental and physical world; they learn that events or stimuli in the physical world can cause mental states that can in turn result in behaviours. Thirdly, children learn that mental states and the physical world are separate and different; for instance people can think about things that are not there. The second phase in the development of a ToM is proposed to be the leap children make to understanding that beliefs can both represent and misrepresent reality, termed here first manifestations of a real theory of mind. The final phase is proposed to account for the understanding that "the mind actively mediates the interpretation of reality" (p. 68). As Muris et al. have explained, this is where children learn that earlier experiences can impact upon current mental states that, in turn, can impact how people experience the social world. The ToM test developed by Steerneman aimed to capture the developmental progression through these three phases.

[1] A Dutch scale developed by Steerneman, et al., not reviewed by the current author.

The first aim of Muris et al.‟s study was to examine validity of the test as a developmental measure; that is, whether the test evidenced increasing success with age. The second was to examine concurrent validity of the test in relation to more common tests of ToM and social development. The new ToM test, so termed by Muris et al., was designed for use with 5 – 12 year olds. It consists of 78 items in the form of stories, vignettes and pictures. Of the 78 items, 29 target precursors to ToM (such as emotion recognition (concept of mental states) and understanding pretence (concept that the real world and mental world can differ), 33 target early ToM (understanding FB), and the final 16 items measure more advanced ToM (second order FB tasks [e.g., people can think about the beliefs of a person about the beliefs of another person] and understanding humour [concept of how people‟s past experience impact the way they think about current events]). For a sample of 70 children, 10 from each age group 5, 6, 7, 8, 9, 10, and 11/12 year olds, the test was found to have good internal consistency. Additionally, scores on the new ToM test correlated strongly with age. With the exception of the Smarties (unexpected contents) test, which evidenced ceiling effects across all age groups, all ToM and social measures used to assess construct validity, correlated strongly with age. For this sample of 5 and 6-year-old children, the Smarties test (M = .9, SD = .3) appeared to be much easier than the Sally-Anne (change of location) test (M = .4, SD = .5). The new ToM test further demonstrated developmental validity evidenced by discriminating well between age groups, and generally conforming to findings of age-related trends in some ToM literature (e.g., Muris et al. identified a sharp increase in ToM ability between ages 6 and 7 years, a similar result was found in the seminal ToM paper with children) (Perner & Wimmer, 1985). The new ToM test also showed construct validity, correlating moderately to strongly with common single item ToM measures and social developmental measures.

In a second study Muris et al. examined test-retest reliability for the measure using a sample of 12 children ranging in age from 5 – 12 years. A significant improvement for children between time 1 and time 2, eight weeks later, was recorded. Using intraclass correlations (range = .80 - .99), test-retest reliability was good. However, the significant difference between test scores for a small sample over eight weeks is somewhat troubling, because in this age group significant developmental change would be unlikely during an eight week time frame. Practice effects may account for this improvement. In this small sample it is possible that a relatively large change in scores (range = 6 – 8 points) of few children (n = 3) contributed to the significant difference between time 1 and time 2. A further four children‟s scores changed moderately (range = 3 - 5 points). The remaining children‟s (n = 5) scores did not change or changed only slightly (range = 0 – 2 points). When taking into account the average differences between age groups for this measure, reported in the first study, and comparing those to the changes in scores across a relatively short period of eight weeks for this study, slightly more than half of the children can be considered to have demonstrated poor test-retest reliability. Test-retest reliability is then not clearly evident from the findings reported in Study 2. An evaluation of test-retest reliability with a larger sample for each age group would be necessary to draw any firm conclusions.

In summary, the new ToM test attempted to capture the developmental progression of concepts involved in the development of ToM. In this respect the test was the first instrument designed to measure any developmental aspects of ToM, rather than simply measure age of false-belief understanding. However, small sample sizes for the assessment of test-retest reliability limit generalisability. This study used only 12 children, one or two from each age between 5 and 12 years, and more than half of these children evidenced moderate to large changes in test score relative to differences in mean test scores between age groups.

A more recent attempt to devise a ToM test battery used fewer items but targeted more specified mental states. To do this Wellman and Liu (2004) first conducted a meta-analysis to compare relative difficulty of a number of related ToM tests, with the aim of devising a scale based on these measures. The results of this meta-analysis indicated that children progress through a series of mental understandings; beginning with an understanding of desires, followed by an understanding of beliefs (including the understanding that people‟s beliefs can differ), followed by an appreciation of ignorance, and only then an understanding of false beliefs. As evident from the meta-analysis, a number of studies have examined differences between a few of these constructs at a time, but this meta-analysis was the first to draw together the literature that has examined these variables.

Based on the findings of the meta-analysis, Wellman and Liu designed a second study to assess the scalability of items designed to measure each developmental ToM progression in young children (aged 3 - 5 years). The term scalability is used here to refer to a tendency of items in the scale to reflect an orderly progression of difficulty, whereby success on a more difficult item generally corresponds with success on all previous, less difficulty, items. The final scale consisted of five items in total, each measuring one mental state; diverse desire (understanding that people can have different desires), diverse belief (understanding that people can have differing beliefs), knowledge access (understanding that seeing leads to knowing and not seeing leads to not knowing), contents false belief (understanding that a person can have belief that does not match reality), and real-apparent emotion (understanding that people can display one emotion but experience a different emotion). In essence this scale measures the concepts tested in the first two phases of Sterneman‟s new ToM test; discussed earlier in this chapter. The main difference between Sterneman‟s and Wellman and Liu‟s measures is that Wellman and Liu‟s scale uses single items to measure a number of discrete concepts (e.g., belief vs. desire), whereas Sterneman‟s new ToM test uses multiple items to measure a general stage of development (e.g., concept of mental states). The new ToM test was also found to be useful for measuring children‟s later ToM development. On the other hand, the scale developed by Wellman and Liu only included a single item, real-apparent emotion, which is believed to be more difficult than FB understanding. An understanding of real-apparent emotion is achieved when a child can appreciate that people do not always behave in accordance with how they feel. This skill is at the foundation of understanding sarcasm, hiding one‟s true feelings, and so forth.

The data indicated that the initial pool of seven items evidenced an increase in difficulty from 95% success on the easiest item (diverse desires) to 32% success on the most difficult item (real-apparent emotion). Two FB items, unexpected contents and change of location, exhibited a statistically non-significant difference (59% vs. 57% success respectively). Given that the type of scalogram analysis (Guttman scale) employed is not able to account for variations in performance for items of similar difficulty, two original items were dropped from the battery. The initial seven items did not scale well if a child passed one but not all items of similar difficulty. The first of the two discarded items was the change of location task. A belief emotion item was also dropped from the original battery because performance on this item (52%) was similar to the FB items. The remaining items proved to fit well to the stringent Guttman Scale. Rasch analysis, for which items of similar difficulty do not pose the same problems, was able to accommodate the two excluded items in a single scale, while confirming these items were similar in difficulty. Wellman and Liu proposed that the similar level of difficulty for the two FB items (unexpected contents and change of location) acted as a check for task equivalence. That is, these tasks, which varied in a number of non-ToM related ways due to differences in presentation or language used, showed similar performance.

To test reproducibility of the scale, Wellman and Liu divided their sample of 75 children into approximate halves, based on the order in which children were tested. Separate analysis of each group produced the same scale, providing good support for the reproducibility of their scale. In total, 60 children in their study (80%) fit the scale exactly. The 15 children who performed in some other pattern were mostly younger children (nine from 25 3-year-olds, four from 25 4-year-olds, and two from 25 5-year-olds). Less predictable performance for younger children is not overly surprising in light of the test-retest literature for false-belief tests that shows that children who perform inconsistently tend to be less verbally mature (Charman & Campbell, 1997); although this was found in a sample of children with learning difficulties. Test-retest analysis did not form part of Wellman and

Liu"s study, but would have been useful in order to be able to assess whether the children who did not perform consistently were generally inconsistent performers due to random responding or inattention. Certainly, for the 3-year-old group, this is not an unreasonable proposition. Single measure items that require a child to attend to a story and answer questions in response to that story can result in unreliable findings for very young children.

Both Wellman and Liu and Steerneman"s attempts to generate a ToM measure that does not rely on individual items for a ToM score, have drawn on a progression of ToM development from the early understandings of desires through to later understanding of more complex social concepts, like misrepresentation of actual belief or emotion. Wellman and Liu and Steerneman"s measures differ, primarily, in the number of test items. They also differ in age range targeted. Wellman and Liu"s scale used more typical presentation methods, using a combination of dolls and pictures, whereas the new ToM test employed only pictures. Both, however, used complex linguistic structures for stories relating to the more advanced test items. These are similar to the story scripts used in common false-belief tests involving change of location. The use of complex linguistic structures, that exceed the linguistic abilities of young children, make it impossible to conclude that younger children lack these social concepts. While it may be the case that children do lack these concepts, it is also likely that younger children lack the linguistic ability to respond intelligibly to these test items. Additionally, younger children may also be limited in working memory and cognitive processing, skills which are required for these tasks. It may certainly be the case that younger children have not developed the necessary concepts in order to pass these more complex items, but in order to conclude this, items need to be developed that are sensitive to the limitations, be they cognitive, memory, or language, of the target age range.

Clinical Samples

Aside from fitting a developmental trend, any test battery for ToM ought to be sensitive to differences in ToM between typically developing children and groups of children that have been identified, in the literature, as displaying a ToM deficit or delay. This includes children with autism (Happé, 1995) and children born deaf to hearing parents (Peterson & Siegal, 2000; Peterson, 2004; Peterson & Siegal, 1995). Both the scale developed by Wellman and Liu and the new ToM test by Steerneman have been assessed for suitability for use with one or both of these clinical populations of children. A summary of these findings follows.

Muris et al.‟s (1999) evaluation of the new ToM test included two studies (Study 3 and Study 4) that validated the measure firstly for use with children with pervasive developmental disorders (considered to be part of the autism spectrum disorders) and then as a tool to discriminate between children with autism and children with other clinical conditions who ought not be delayed in ToM. The results of their Study 3 found high internal consistency and interrater reliability for the measure in this clinical sample of 10 children with an autism spectrum disorder. The sample of children was found to have average intelligence and were, therefore, likely to be at the high functioning end of the spectrum. They ranged in age from 6 years 9 months to 12 years 9 months. Scores on the ToM test ranged from 32 to 75 points out of a possible 78. No other analyses were reported for this sample. The final study, Study 4, compared performance of 20 high functioning children on the autism spectrum (diagnosis of pervasive developmental disorder not otherwise specified [PDDNOS] or autism) with performance of 14 children with ADHD and 18 children with an anxiety disorder. Again internal consistency of the measure was good. Children with autism gave significantly poorer performances on the new ToM test than did those with other clinical diagnoses, despite having similar verbal IQs. The groups did, however, differ on performance IQ, as measured by the WISC-R; where, on average, the children with an anxiety disorder had higher scores than the other two groups. The highest correlation between ToM performance and the IQ measures was for verbal IQ ($r = .61$), although the correlation between ToM and performance IQ ($r = .45$) was also significant. Regression analysis found that a diagnosis of autism accounted for 48% of variance for ToM; where children with autism had lower ToM scores. Additionally, verbal IQ and then age explained further variation in ToM (10% and 6% respectively). Finally a diagnosis of PDDNOS explained a further 4% of variance. PDDNOS is considered to be symptomatically less severe but similar to autism. These results indicate, as highlighted by the authors, that the test has good discriminant validity, in that it is sensitive to the ToM deficit in this population and can discriminate between children who have differing degrees of symptom severity.

In summary, appropriateness of the new ToM test for use with children with an autism spectrum disorder population was not addressed by Muris et al. (1999). The test uses items similar in language and story complexity to common FB tasks. Although the new ToM test showed good internal reliability, and performance on the test by children with autism was as predicted, the test may underestimate ToM ability in this group. Scores on the test were most correlated with verbal IQ, which would seem to support the proposition that performance on the test is limited by language ability.

Wellman and Liu"s ToM scale has also been assessed for use with children with clinical diagnosis, children with autism and late signing children born deaf. Peterson, Wellman and Liu (2005) examined whether the stages of deaf children"s ToM development were similar to those found earlier in typically developing children. Peterson et al. included three clinical groups and a control group. Their sample of 145 children included: children born deaf in a family with a deaf parent who had no delay to learning sign language - the parent is referred to as a „native signer"; the group is included to control for aspects of deafness not related to developing a ToM - (n = 11, mean age = 10.67 years, SD = 1.83); children born deaf into a family with no native signers who experienced a delay in learning sign language (n = 36, mean age = 10.01 years, SD = 2.46); children with autism (n = 36, mean age = 19.32 years, SD = 1.88); and typically developing children (n = 62, mean age = 4.5 years, SD = .59). Language ability was controlled for in the clinical groups in order to include only those children whose language skills were sufficient for the tasks. For the group of children with autism, a minimum verbal mental age (VMA) of 4 years was required; mean

VMA for the group was 7 years 10 months as measured by the Peabody Picture Vocabulary scale. For children born deaf, a minimum teacher rating of 2 on a 6-point scale was required (indicating a "somewhat smaller than average vocabulary, but adequate for everyday communication" [p.505]). Typically developing children were not tested for verbal ability.

Consistent with the literature, Peterson et al."s findings supported an account of ToM delay in deaf children born to hearing parents and a ToM difference in children with autism. Peterson et al. found that, on average, children born deaf to signing parents outperformed children born deaf to non-signing parents. Performance on the scale in the former group was high; 100% passed the knowledge access item, 91% passed the diverse beliefs item, 82 % passed the diverse desires item, 82% passed the FB item, and 54% of these children passed the hidden emotion item, compared with children born deaf to non-signing parents, where 92%, 92%, 53%, 33%, and 24% passed each of these five items, respectively. As noted by Peterson et al. the finding of better FB test performance of the children born to native singing parents were in-line with much of the published literature in this field. Children with autism showed a slightly different pattern of performance on the scale, with 86%, 86%, 75%, 47%, and 64%. Order of items in the scale reflected the proposed order in which these abilities are acquired in typical development. The data, therefore, showed that children generally performed better on the measures corresponding to developmentally earlier concepts and performance decreased for subsequent items in the scale that corresponded to later stages of development. The only exception was that the FB item appeared to be particularly difficult for children with autism compared with the developmentally more advanced hidden emotion item. Guttman analysis of the current data from typically developing children and deaf children fit well with the results reported by Wellman and Liu (2004). The same analyses found that children with autism did not perform similarly on the scale. For this group the order in which ToM concepts were acquired fit best with a scale for which the last two items (FB and hidden emotion) were reversed. No scaleogram analysis was possible for the group of children born deaf to native signing parents, due to the small sample size of this group. Nine of these 11 children, however, performed in a way that fit the scale exactly.

Reliability and validity of the scale were better addressed in the Peterson et al. study than in the Wellman and Liu study. Peterson et al. reported percentage of children who passed control questions, where applicable, which is useful for determining whether children have a general difficulty with the item due to task complexity or, more specifically, if the difficulty is related to the ToM aspect of the task. Peterson et al. considered that passing the control question for an item but not passing the ToM test question indicated that the task was not too complex. A pattern, similar for each group of children, of decreased performance for control questions as item difficulty increased supports the notion that task complexity, and not merely ToM difficulty, increase across tasks. Nonetheless, the lowest percentage of children who passed the control question was 70%, indicating that, for most children, the tasks were not excessively complex. It is likely that task complexity does contribute to poor ToM performance but that it does not obscure performance entirely.

To address issues of suitability of any ToM measure it is necessary to consider the effect of language ability on ToM test performance. Although language has been found to be the most important factor in ToM performance in children with autism (Fisher, Happé, & Dunn, 2005; Happé, 1995), it is unclear as to whether language ability is a performance inhibiting/prohibiting factor or whether language and ToM are developmentally entwined. In typically developing children it appears that language enables ToM development. Longitudinal research has found that early language predicted later ToM but that the reverse relationship was not evident (Astington & Jenkins, 1999). Children with autism develop very differently from typically developing children. Differences in experience begin with sensory differences (Bogdashina, 2005) and extend to differences in social interactions (Carpendale & Lewis, 2004; Shanker, 2004). These fundamental differences must certainly contribute to any development of social functioning. Sensory processing is fundamental to any learning, be it about one‟s own body, the physical world, or the social world. If children with autism suffer from sensory overload or under-load, as discussed by Bogdashina, these differences will have a ripple effect for development across the board. Shanker proposed that an absence of triadic interactions impedes social development in autism. That is, children learn about the social world, and indeed the minds of other people, through shared emotional and environmental experiences. For whatever reason, these are not common place in the lives of children with autism. This may be because children with autism lack a social imperative, that they find the social world overwhelming, or for some other reason. A lack of these experiences makes learning about social concepts and other people‟s minds difficult.

Construct Validity: Criticisms from a Social Perspective

Although some research has examined construct validity of ToM test batteries, in relation to common false belief tests and to some measures of social functioning (Muris et al., 1999), little is known about how ToM enables or disables, given its presence or absence, social functioning in children and individuals with autism. The ToM literature has examined the influence of social stimuli on the development of ToM such as, mother‟s use of mental state terms (Slaughter, Peterson, & Mackintosh, 2007), number of siblings (Perner, Ruffman, & Leekam, 1994; Ruffman, Perner, & Parkin, 1999), mother‟s level of education and disciplinary style (Pears & Moses, 2003), and overall environment (Hughes, et al., 2005). The findings of this body of literature indicate the importance of children‟s social experiences in developing ToM. Carpendale and Lewis (2004) argued that any model of ToM needs to account for these social environmental influences. Despite a relatively large literature examining the impact of the social world on the development of ToM, a much smaller literature examines the impact of ToM on the social world. This literature concerns itself with the construct validity of common ToM tasks; i.e., whether ToM as measured by FB and similar tasks, is useful for predicting social functioning. Consideration of this literature follows.

Theory of Mind in Autism

The first criticism of ToM came from a phenomenological analysis of the autobiographical writings of 10 high-functioning individuals with autism. Williams (2004) summarised the nature of the social difficulties experienced by these individuals as being rooted in the flexible nature of people. That is, that regardless of how many rules for social interaction people with autism can learn, the problem they experience in actual interactions is that they are unable to apply these rules effectively to the ever changing behaviours of people. In other words, people do not conform to the learned rules of social interactions. Williams proposed that any account of ToM that posits an automatic process that, once a child has developed a set of concepts (rules), is engaged and functions to mediate social interactions, disregards the mechanisms necessarily involved in applying these rules.

A ToM model must then account for the mechanism that learns the rules and also a mechanism that allows for these rules to be seamlessly accessed, applied, and adjusted during complex social interactions. Williams argued that neither of the prominent and central theories of ToM are able to explain this. Instead, Williams proposed that social interaction can be thought of as ranging from beginner to expert. In summarising a model proposed by Dreyfus and Dreyfus (1991) Williams explained that experts do not need to revert to theories in their every day interactions within their field of expertise. Instead, experts simply know how to respond to a particular common event and only need to access rules to address unusual circumstances. Beginners on the other hand need to refer to rules to cope with relatively novel situations. Seemingly, this account places emphasis on rehearsal and practice, whereby these factors determine the type of cognition engaged in any situation. This account is intuitive and difficult to argue against. A strength of the theory is that it can easily account for the findings of the effect on social environmental experiences in the development of ToM proficiency. It is also able to account or findings of ToM delays in children who have impoverished social experiences (e.g., children with autism and children born deaf to hearing parents). Impoverished experiences result in delayed development of ToM and a disability of the same mechanisms by which ToM is acquired (engagement in the social world) continues to impede effective application of the learned rules.

Williams‟ article also highlighted a means by which social engagement in autism is impaired. The accounts of what social interaction is like for people with high-functioning autism and Aspergers Syndrome highlighted the impact of sensory processing difficulties on the ability to „read‟ people. As Temple Grandin, a well know woman with high-functioning autism, has described

> "Since people with autism require much more time than others to shift their attention between auditory and visual stimuli, they find it more difficult to follow rapidly changing, complex social interactions" (Williams, 2004; p. 713)

Similarly Williams quotes Darius, who has Aspergers Syndrome

> "I still have problems in groups where there is too much information going on, or where the information flow is too fast. ...Consequently I miss a lot of vital information needed to interpret social messages. There is no such thing as adequate delayed social interaction. One is either quick enough to keep up, or one is weird and socially disabled" (Williams; p. 713).

Williams suggested that the reported sensory disturbances experienced by people on the autism spectrum interrupt emotional engagement with other people; a process that requires awareness and appropriate response from both parties to physical and verbal information offered by the other. The experiences of people who are able to communicate what it is like to have autism and how the symptoms of autism affect their daily interactions, must surely be taken into account when forming a theory of how a missing ToM manifests itself in the lives of people with autism. As concluded by Williams, people with autism do not lack a theory of other people‟s minds; in fact this is precisely what they develop. Instead, the theory developed is not sufficient to ensure successful social interaction.

The accounts of people with autism were published autobiographies and Williams provided no indication of how these people perform on tasks assessing ToM. One would assume, given these are all high-functioning adults, who are for the most part University educated, some holding titles of Professor, that they would perform well on typical FB type tests, which are designed for use with children and have been found to be passed by children with autism who have a verbal mental age above approximately 9 years (Happé, 1995). This is mentioned here, not to assess the ToM of these adults, but instead by way of indicating that passing ToM tests is not a sufficient indicator of social ability. However, in order to generalise the autobiographical accounts to the general population, a systematic investigation of the issues and themes raised is required. This self-reported data set provides a good starting ground for any such investigations.

Theory of Mind in Very Young Typically Developing Children

Reddy and Morris (2004) have extended the debate raised by Williams about whether ToM was a useful social construct to young children who, according to the literature do not yet possess a functional ToM. The authors firstly take issue with the primary assumption underlying theories of ToM, that is, "that minds represent, and that interacting with minds requires, and is therefore constrained by, an understanding of representation. Such understanding is therefore seen as a prerequisite for appropriate engagement with other minds." (p. 648). Reddy and Morris have presented evidence, which they say supports the idea that young children actively engage in mental state activities with other"s in their social environments long before they pass tests of false belief and are deemed capable of representing and theorising about other minds.

Reddy and Morris presented a range of findings in order to cast doubt on the claim that children younger than 4 years lack an understanding of mind. The first form of findings were from both laboratory studies and naturalistic observations of children as young as 2 years old deliberately deceiving others. Findings of early deception have, according to Reddy and Morris been dismissed in the greater ToM literature as rote learnt verbal strategies developed in order to avoid punishment. Summarising this body of the literature"s findings, the authors reviewed the motives behind early acts of deception and reported that these deceptions are used by children to save face, for trickery, and also for fun and not simply to avoid punishment. The authors dismiss claims that rote learnt behaviours do not constitute real knowledge, citing the variety and complexity of lies that children use, which they say cannot simply constitute rote learnt strategy. While these behaviours are real and occur in children who have some social knowledge, this degree of understanding of social knowledge differs from a more complex level of understanding of the social system akin to that of adults, which is likely not present in very young children. A rudimentary understanding may well be present in young children, which grows and changes as the child‟s cognitive resources expand with age.

The second set of findings presented by the authors demonstrates that children act to correct ignorance of other people. They argued that children without a ToM would not recognise ignorance, nor think to correct ignorance. The evidence presented takes the form of children aged 2.5 to 3 years using information relating to the presence or absence of their mothers to request a toy that was out of reach. Specifically, children whose mothers were absent provided more information about the location of the toy than did children whose mothers were present when the toy was relocated. Similar research assessing non-verbal information given by children found that even 18-month-old children gave more non-verbal information indicating the location of the toy when their mother was not present while the toy was relocated. Reddy and Morris explained that the earlier conclusions, drawn by others, that young children are not able to recognise and compensate for ignorance, comes from research involving a hide and seek game, which they argued was not age appropriate. Children were able to hide but were not able to do so in a way that took into account the seeker‟s knowledge. It is perhaps the case that hide and seek games, which require children to follow complex rules and procedures (close your eyes and keep them closed, don‟t hide somewhere predictable, etc), obscure the fundamental understanding of young children because other cognitive mechanisms are not yet in place to help the child manage the game requirements.

Certainly, young children often attempt, at a very early age, to hide from their parents actions that they know are frowned upon.

Another form of mentalising discussed by Reddy and Morris is the way in which children act to repair miscommunication. The authors suggested that, if we are to subscribe to the notion that children only develop an understanding of misrepresentation after they can pass a test of FB around the age of 4, we must conclude that children who have not yet developed this concept are also not able to recognise misunderstanding or misperception. The authors then presented Golinkoff‟s findings that children as young as 14 months repair other people‟s misunderstanding of their communicative attempts. This is not necessarily convincing evidence of an early ToM in young children, however. Repairing miscommunication can be achieved without any need to access another person‟s mental state. For example, if a child requests something and the parent gives the child something else, the child does not need to understand that his/her parent misunderstood in order to request again. Indeed Shatz and O‟Reilly (1990), as highlighted by Reddy and Morris, proposed that communicative repairs are a strategic behaviour and not reflective of an understanding of mind. It could, however, be argued that if a child clarified his/her earlier attempt by providing more or different information, then this indicates an understanding, even if limited, of the other person‟s mind. In this scenario the child adjusts his/her behaviour or request in order to provide the necessary information in order for the parent to understand; thereby recognising that „understanding‟, which is a mental state, is not present in his/her parent.

The final demonstration of young children‟s concept of mind was made in relation to declarative pointing. The authors argued that a child‟s declarative pointing is a communicative act, rather than a behaviour. Reddy and Morris proposed that to understand these acts as well as the aforementioned evidence of children‟s interactions with other‟s minds, it is important to separate oneself from the popular view of children‟s theories about other people‟s minds, thoughts, and so on. Instead the authors believed it is imperative to view interactions between people, be they children or adults, as acts of engaging rather than behaving. They drew a distinction between people treating other people as „objects‟, which they need to try to understand, and people engaging with another „person‟. From the perspective of Reddy and Morris, these engagements involve an emotional exchange in which people experience the other person rather than observe them. In this way a child is not seen as an observer of the people with whom s/he interacts. Instead, the child is a participant in these interactions. From these shared interactions the child is thought to learn about people.

This perspective differs from the most popular theories about ToM in that it does not separate the mind from the body. One of the most common ToM accounts proposes that children gain, in their early life, an understanding that other people have minds, which are separate from their own, and that their desires coupled with their beliefs govern their physical behaviour. The physical is distinctly separate but governed by the mental. According to this theory a child is like a scientist exploring a new phenomenon. The child is thought to transition through a set of understandings, the first of which depicts people as comprising a set of behaviours, after which the child learns that behaviours are governed by desires, which are eventually followed by the understanding that desires combined with beliefs govern behaviour (Wellman and Liu, 2004). This staggered developmental view, which locates the child as an observer of people, does to a large extent ignore the types of naturalistic interactions a child has with those people in his/her social environment. In essence it removes the social aspect of socialising and replaces it with experimentation; experimentation that leads to knowledge and understanding. Reddy and Morris appeared to argue that it is in fact this social „experience‟, rather than observation, that leads to knowledge and understanding. They proposed that thinking of social development in this way explains children‟s early communication and interactions with others in a much more natural and appropriate way.

To conclude, most literature that contests the construct validity of ToM research and its measures takes issue with the tendency of the ToM literature to position people as observing and explaining others‟ actions rather than as interacting with others. This literature argues that the populous ToM literature routinely dismisses children‟s early signs of interacting intentionally with other minds by, instead, formulating these interactions as rote learnt and strategic. Reconciling this social account with the predominantly behaviourist accounts of ToM is, however, not entirely problematic. Concessions and adjustments to theories would be necessary, but it is possible to accommodate both interacting and learning in an account of children‟s social development. As is evident from autism, without the ability to interact effectively the learned knowledge is cumbersome to come by; and when it is acquired it is not effectively used. Similarly, without the ability to learn rules to govern interaction, these interactions would not develop and change, which they clearly do. Indeed the most rational theory would define social interactions as the means by which children learn about the social world.

Conclusions

A review of the literature examining the reliability and validity of ToM measures has highlighted the importance of future research to develop better measures and to consider alternative points of view in order to advance Psychologists‟ understanding of social development.

FB tests, ToM test batteries, and ToM scales all have their roots set in a behaviourist framework. The tests assess children‟s abilities to observe, predict and explain other‟s behaviour, rather than to interact with other people‟s minds in a sensible and natural way. Aside from this, the measures can also artificially disguise understanding and knowledge of others, mainly because young children have limited cognitive resources. If, however, we are to accept that the tests provide an accurate picture of children‟s abilities, it is the case that these are not necessarily related to children‟s social abilities (Astington, 2003). The question then is whether current measures of ToM are useful at all.

The most convincing explanation for social delays and deficits in autism comes from research that accounts for and incorporates evidence of sensory difficulties in autism. As discussed by Williams (2004), social interactions are stimulus rich, which poses problems for people with sensory and processing difficulties when interacting with and learning about social beings. Once people with autism have learned rules that govern human behaviour these do not help them navigate social interactions. Having a ToM, as tested by various ToM tests, does not make the social world of people with autism easier. Arguably their difficulty lies not in poorer understanding but in effectively interacting. Similarly, young children successfully interact without having an apparent understanding of minds. In this way a young child without ToM is nothing like a person with autism without a ToM and is certainly more appropriately social than a person with autism who does have a ToM.

Nevertheless, a ToM test can be a potentially important tool, if limitations are acknowledged. There are research situations in which naturalistic observations or parent interviews are not informative. Neurological research, for instance, which examines brain damage, be it from an accident or an underlying physical condition, needs to use tests in some instances that can stimulate brain function and in conjunction with fMRI help to map the brain (Saxe, 2006). Nonetheless, ToM tests have barely changed and only slightly evolved since their invention in 1983. Twenty six years of research, which has for the most part ignored other perspectives, has not made any radical advances in helping children with autism to develop social skill or achieved any radical insights into the social development of children. Many children and adults suffer from social impairments and advances in understanding of the complex phenomenon of social development must therefore take priority over scientific rigidity and inflexibility. On a closing note, theories are only useful in so far as they provide the scientific community with the stepping stones to developing tools to help and make a difference for society.

Rationale for Study of Presentation Method

The current chapter has reported published accounts of reliability and validity for a range of ToM tests. The literature, as critically evaluated and summarised here, has not included a direct experimental assessment of the contribution of minor or major differences between tasks to the difficulty of the tasks for young children. The next chapter presents such a comparison by way of illustrating how changes to tasks can increase or decrease task difficulty. The literature presented in the current chapter gives some indication that presentation method can affect task difficulty and task reliability. Therefore, the major aim of the next study reported in this thesis was to assess whether the ToM battery from Chapter 4 was unduly difficult for 4 and 6-year-old children because of the presentation method used.

CHAPTER 6 - Measuring Theory of Mind with False Belief Tests: Comparing 2D and 3D Presentation Modes

Abstract

Standard False Belief (FB) tests can be presented using either two dimensional (2D) or three dimensional (3D) methods. Children‟s variability in performance on FB tests in response to method of presentation was examined. Thirty children (M = 4.5 years, SD = 4 months) each completed a series of 12 false belief tests, six presented using props (3D) and six presented using cartoon strips (2D). Order of presentation of the tests was manipulated according to dimensionality and difficulty of the FB test. Children were more likely to pass 3D FB tests than the 2D tests, but no significant order effects for presentation were found. However, patterns of increased success on more difficult test items were evident when simpler tasks were presented first. Implications for ToM research are discussed.

Measuring Theory of Mind with False Belief Tests: Comparing 2D and 3D Presentation Modes

Theory of mind (ToM) has been assessed using False Belief (FB) tests since Wimmer and Perner (1983) first introduced the concept for use with children. These tests have been found to reliably detect above-chance performance in preschool-aged children, across a range of tasks and ability levels (Hughes et al., 2000; Wellman, Cross, & Watson, 2001). While researchers have generally agreed that ToM begins to emerge in 4-year-old children, there have been reports of earlier success; that is, researchers using FB tests have reported above-chance performance on tasks by children younger than 4 years, as summarised by Wellman, et al. (2001). Earlier success may be a result of variations in the presentation of tasks. Moreover, standardised FB test batteries do not exist, leaving researchers who use the FB tests to create their own tests, within „standardised‟ parameters. Although an extensive meta-analysis suggests that on the whole this does not create problems for the reliability of the fundamental components of the tasks (Wellman et al., 2001), variations across different studies can lead to differences in test success rates, conclusions drawn, potentially adding noise to data sets, and creating uncertainty in relation to the definition of the construct.

Aside from issues of reliability across different types of FB tests, some have questioned the ability of the FB task to measure young children‟s ToM accurately. The standard false belief test has two widely used forms; the Sally-Anne task, which Baron-Cohen, Leslie and Frith (1985) adapted from the Maxi chocolate task (Wimmer & Perner, 1983), and the unexpected contents/identity task (Perner, Leekam, & Wimmer, 1987). Bloom and German (2000) summarised evidence of ToM in younger populations, concluding that these standard FB task place substantial processing demands on preschool-aged children who have inefficient processing capacities. They concluded that the FB task is not able to detect early ToM.

A task that usually detects ToM in children at an earlier age than the standard FB task is the deception task, in which success is measured by the child‟s ability to create a deception (Chandler, Fritz, & Hala, 1989). Deception tasks usually involve the child erasing a character‟s foot prints or creating a false set of foot prints, in order to conceal the true location of an object (Chandler et al., 1989). The rationale for deception tasks is that children need some understanding of other people‟s minds in order to deceive, that is, to create a false belief. Deception tasks differ from false belief tasks in a number of important ways. Firstly, deception tasks actively involve the child, in contrast to the observer status of the child for FB tasks. Secondly, a deception task does not rely on a child‟s ability to respond to complex language formations as is the case for the FB task (for a comprehensive review of language factors see Milligan, Astington, & Dack, 2007). Instead, it requires that the child be able to follow a heuristic (e.g., „clues give away the position of an object, so if I leave no clues they cannot find it‟). Furthermore, deception tasks usually involve some form of training scenario that prepares the child for the subsequent task. For instance in Chandler et al.‟s experiment children were first asked to try and find the object using clues left behind by the person who hid the object. Children were then asked to help hide the object so that a person, who had left the room, would not be able to find it. These experimental procedures attune children to deception strategies. FB tasks can also prime children for the test question by asking a number of control questions before presenting the test question. These questions assess memory of the story and by doing so potentially alert the child to important factors like the change of location and naivety of a character. Arguably these questions do not give the child the tools to answer the question in the way that deception tasks provide deception strategies. Additionally, some researchers ask these questions after the test question and therefore provide no priming.

Aside from differences between FB tasks and deception tasks, the Sally-Anne task and the unexpected identity task also differ in complexity in the degree of linguistic demands placed on the child (Milligan et al., 2007). Where the Sally-Anne task requires that the child follow a story in which a series of events takes place, the unexpected identity task makes no such cognitive demand. Both tasks require the child to draw on a representation of another person‟s false belief, but the unexpected identity task does this with what is arguably less cognitive workload. However, in a meta-analysis of the relationship between ToM and language, Milligan et al. (2007) found that the strength of the relationship between language ability and individual FB tasks, which varied in linguistic complexity, did not differ significantly, despite a main effect across four types of FB tasks. The authors also compared the relationship between FB task performance on all types of FB tasks combined together, and different types of language measures utilised in the literature, finding that general language measures were more strongly related to ToM performance than were receptive language measures. Milligan et al. (2007) did not report an analysis of differences between the relationships of specific FB tasks for each of the different language measures. However, such an analysis could better detect differences in the relationships between FB tasks and language tasks, given that FB task and language measures both vary in their degree of complexity.

In a meta-analysis examining age of onset for ToM, Wellman et al. (2001) reported evidence that children‟s ToM does, in fact, undergo a conceptual change in preschool years and that there is uncertainty as to when this change occurs. Wellman et al. nonetheless found that some task manipulations like "framing the task in terms of explicit deception or trickery, involving the child in actively making key transformations, and highlighting the salience of the protagonist‟s mental state or reducing the salience of the contrasting real-world state of affairs" (p. 672) can improve young children‟s performance on ToM tasks. Leaving aside, for the moment, the question of the nature of ToM, reported variations in ToM ability due to differences between tasks create uncertainty in the literature about the usefulness of FB tests. Although nonstandardised FB tasks continue to be used in developmental research (Bright-Paul, Jarrold, & Wright, 2008; Kobayashi, Glover, & Temple, 2008; Ontai & Thompson, 2008; Sabbagh & Seamans, 2008; Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008), arguably what is required is a standardised task, upon which the research community can agree, to map children‟s development.

This study expands on research presented in Chapter 5, which reported poorer than expected ToM performance in 4-6 year old children when tasks were presented using cartoon strips in place of puppets or props. The study utilised standard FB tasks, both unexpected identity and change of location tasks, presented as cartoon strips. In Chapter 5 it was proposed that increased task difficulty may have stemmed from the cartoon strip presentation method. Outside the field of ToM research, studies have examined participant performance on content identical tasks presented with different mediums. For example, participants who used a computer to complete a 15 minute workload task, which measured participants perceived workload when completing a number of tasks, experienced greater workload than participants who completed an identical paper and pencil version of the task (Noyes, Garland, & Robbins, 2004). Additionally, developmental psychology proposes that learning is hierarchical, where people build on previous knowledge as they learn new skills. For instance, children have knowledge of objects before they learn that symbols can represent real objects or environments (Marzolf & DeLoache, 1994). This raises the question of whether the visual method of presentation of standard FB tasks to test ToM in pre-school aged children may impact on pass rates for these tests.

Based on the increased task difficulty for cartoon strip presentation of standard FB tasks reported by in Chapter 5 and reported findings of performance differences in relation to test mode, we hypothesise that 4-year-old children will perform better on FB tasks presented with dolls and props (3D) than when tasks of equal construction are presented using cartoon strips (2D).

Method

Participants

Data from 30 children from South Australian pre-schools were included in the analyses (12 males and 18 females, M = 4 years 6 mths, SD = 4 mths). Children (n=11) excluded from data analysis included: one boy diagnosed with Autism; three children not proficient in English; one girl diagnosed with a language disorder; three boys who did not pass the ToM test control questions; and three boys who did not cooperate with researcher instructions. Recruitment was conducted through preschools. The study had human research ethical approval. Consent for participation was obtained from parents or carers and preschool directors. Children were from middle to high socioeconomic groups, all attending preschools in metropolitan Adelaide.

Materials

Receptive Language Measure

The receptive language component of the Communication domain of the Battelle Developmental Inventory-2 (BDI-2) (Newborg, 2005) was used to control for language ability in children. It measures developmental achievement, an important factor in comprehension of ToM stories. This measure was used instead of a vocabulary-based assessment of language, because findings of an earlier study suggested its relationship to ToM is more robust in this age cohort (see Chapter 5). Furthermore, the BDI-2 was not already used by preschools involved in the research project for any assessment of children for school readiness and it was therefore not vulnerable to test-retest problems. It was administered according to BDI-2 instructions.

Theory of Mind

ToM was measured using a series of six 2D and six 3D traditional FB tasks.

The results presented in Chapter 5 indicated that children had particular difficulty with ToM tasks presented in 2D mode. The same test 2D test battery, as that reported in Chapter 5, was used here, with stories presented with cartoon strips, rather than with puppets or dolls; two stories involved a person changing location and two involved an object changing location, similar in nature to traditional Sally-Anne tests (Wimmer & Perner, 1983). The final two test items did not involve a location change but instead used a false contents test. These were also presented as picture cards rather than 3D objects.

To assess ToM ability for 3D presentation mode, presentations used dolls and objects to act out stories. Order of presentation and type of story (person/object changing location or false contents/object) were as for the 2D test battery.

A repeated measures design assessed whether ToM test presentation mode influences test outcomes. Each child completed a total of 12 ToM test items, six 2D and six 3D. To assess order effects, 16 children completed the 2D battery first and 14 completed the 3D battery first. Similarly, 16 of the children completed the ToM test batteries in the order from most difficult ToM questions through to the least difficult and 14 children completed the reverse order. Story scripts for the 2D battery were taken from the study presented in Chapter 5. Story script outlines and examples are presented in Table 1.

Table 1

Theory of Mind Story Script Outlines

Task type	Component	Outline	Example
Person change location	Story	X and Y were (action) in (location).	Ben and Sarah were playing together in the playground.
		(Event) and X left (location) to (action).	Sarah heard the ice cream truck and went to buy and ice cream.
		While X was gone, Y (event) and went to (location) to (action).	While she was gone, Ben saw a cat stuck in a tree and climbed the tree to get the cat down.

		X has finished (action), and wants to (action) Y.	Sarah bought an ice cream for Ben and one for herself. She wants to give Ben his ice cream.
	ToM test Q	Where will X look for Y?	Where will she look for Ben?
	Control Qs	Where is Y really?	Where is Ben really?
		Where was Y before s/he went (location)?	Where was Ben before he climbed the tree?
		Did X see Y go to (location)?	Did Sarah see Ben climbing the tree?
Object change location	Story	This is X and this is Y	This is Lisa and this is Paul.
		X has (object) and puts it (location)	Lisa puts her ball in a basket.
		X (exits scene with reason)	Lisa goes to get some lunch.
		While s/he is gone, Y (action)	While she is gone, Paul wants to play catch so he takes Lisa"s ball
		When (action is concluded) (object is moved)	When he is finished he puts it in a box.
		X returns and wants (object)	Lisa comes back and wants to get her ball.
	ToM test Q	Where will X look for his/her (object)	Where will she look for her ball?
	Control Qs	Where is (object) really?	Where is the ball really?
		Where was the (object) before Y (action)?	Where was the ball before Paul played catch?
		Did X see Y move (object)?	Did Lisa see Paul move the ball?
False object/ contents	Story	Introduce object and ask what child thinks it is or what is inside.	Look at this picture. What do you think this is this?
		Offer to reveal true nature.	Let"s turn it around and see.
		Reveal true identity	It"s a bird.
		Establish another child"s naivety	(Child"s friend"s name) is

		coming in next and s/he hasn‟t seen this picture before.
ToM test Q	Assess ability to express understanding of FB	If I ask (Child‟s friend‟s name) what this is, what will s/he say it is?
Control Q	Assess child‟s memory of true identity	What is it really?
	Assess child‟s memory for own earlier FB	What did you say it was before you saw that it was a bird?
	Assess child‟s memory of other‟s naivety	Has (Child‟s friend‟s name) seen that this is really a bird?

Procedure

All children were tested at their preschool in a quiet, distraction-free environment. Testing times varied between mid-late morning and early-mid afternoon. All children completed the full test battery. Testing took up to 40 minutes per child, and was completed in a single session. The order of test administration was Communication domain of the BDI-2 followed by ToM test batteries.

Results

Sample Descriptives

The mean receptive language ability of children participating in this study was above average at 5 years 4 mths (SD = 9 mths, Range 4 years 1 mth – 7 years 4 mths) equating to the 82nd percentile within their cohort (SD = 20 percentile points, Range 25th – 99th percentile).

Difficulty of 2D vs. 3D ToM Tests

Table 2 displays mean ToM achievement scores and SD for ToM test batteries for: presentation mode (2D vs. 3D); order of administration for presentation mode (2D first or 3D first); and order of test item administration (unexpected identity items first or change of person items first). A repeated measures ANOVA demonstrated a main effect for presentation mode, with children performing better on the 3D test battery [$F(1,26) = 8.32$, $p < .01$, $\eta^2 = .24$]. There were no interaction effects so that differences in performance were not due to order effects; neither presentation mode order [$F(1,26) = 3.43$, $p = .075$, $\eta^2 = .12$] nor test item administration order [$F(1,26) = .086$, $p = .77$, $\eta^2 = .003$]. However, presentation mode order approached significance and the implications of this will be discussed later in this report. Small samples for these analyses have limited power to detect small effects, as indicated by effect sizes. Nonetheless, the findings support the hypothesis that children perform better on FB tasks presented in 3D when compared to tasks presented in 2D.

Table 2

Mean ToM Battery Scores Relative to Dimension Mode Order of Presentation.

Test battery	N	M Score (SD)	Presentation Mode Order	n	M Score (SD)	Test item order	n	M Score (SD)
ToM 2D	30	2.90 (1.77)	2D first	16	2.5 (1.75)	Unexp. identity 1st	9	2.56 (2.01)
						Change of person 1st	7	2.43 (1.51)
			3D first	14	2.89 (1.54)	Unexp. identity 1st	5	4.20 (1.92)
						Change of person 1st	9	2.89 (1.54)
ToM 3D	30	3.70 (1.82)	2D first	16	3.75 (1.98)	Unexp. identity 1st	9	4.00 (2.18)
						Change of person 1st	7	3.43 (1.81)
			3D first	14	3.64 (1.69)	Unexp. identity 1st	5	4.40 (1.82)
						Change of person 1st	9	3.22 (1.56)

Additionally, to illustrate the pattern of performance for the type of false belief task in relation to presentation mode, mean scores have been presented in Figure 2. The graph shows an increased pattern of difficulty across the FB task, where unexpected identity tasks are easiest for children and task in which a person changes location are the most difficult.

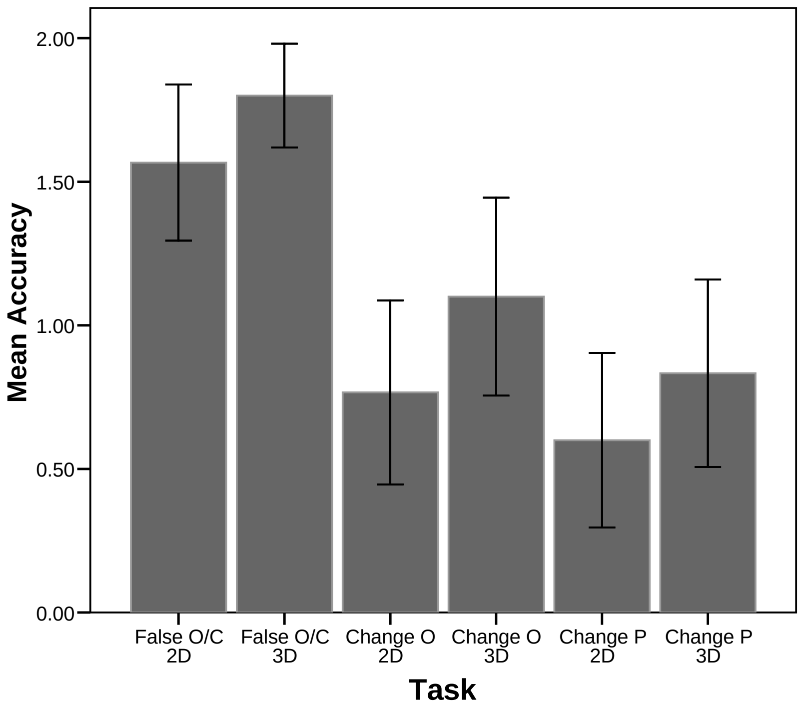

Figure 1. Mean number of correct responses for test items

Note: O = Object C = Content P = Person. Two test items for each task; maximum score = 2. Error bars represent 90% confidence interval

Discussion

Results illustrate how a difference in the way a FB task is presented can impact task performance. This has confirmed the need for standardised testing procedures in ToM research. Children in this study performed below chance on tasks presented using a 2D cartoon strip compared to tasks presented in 3D using dolls and props. Where previous research has demonstrated conditions under which ToM task difficulty is reduced (for a comprehensive review see Wellman, et al. (2001)), this study has demonstrated that presentation mode can increase standard task difficulty. In contrast, although Wellman et al.*'s* (2001) meta-analysis found that some task manipulations can impact ToM performance, they did not find a main effect for what they called "nature of the target object" - that is whether tasks are presented with pictures, video, or objects - although the main effect approached significance. The present findings of a marked difference between performance on FB tests for 2D versus 3D presentation is difficult to compare to the aforementioned findings, given that neither effect size for the analysis nor differences between each of the three conditions were reported.

In addition to task presentation mode differences for the FB test batteries, a clear pattern of variation in difficulty also exists for test items, similar to that reported in Chapter 5. Using the same 2D stimuli as here, Chapter 5 reported a difference in task difficulty between these two commonly used variants of the FB task in both 4 and 6-year-old typically developing children. A similar pattern of performance across these tasks was found in the current study; that is, children performed better on the unexpected identity/contents items than they did on the two test items involving change of location.

In contrast, others have reported no such differences in degree of difficulty for types of FB tasks (Milligan et al., 2007; Wellman et al., 2001). The analyses employed by Milligan et al. did not directly test for these differences. Instead, the claim resulted from a comparison of the relationships between each type of FB task and language ability, as assessed by a range of language ability measures. At the very least, Milligan et al.‟s analysis disregarded differences between the linguistic demands of the FB tasks in relation to the specific facets of language measured by each type of language task. Furthermore, most analyses did not account for changes in the relationship between language and FB tasks in relation to age; an artefact of the studies included in the meta-analysis. Participants in these studies were under the age of 7 years but with actual ages unspecified. Most analyses did not control for age and an analysis of those that did control for age found decreased strength of the relationship between ToM measures and the various types of language measures. Therefore, Milligan et al.‟s analyses cannot speak to differences in the relationship between language and ToM, across age groups, should such differences exist (e.g. differences noted in Chapter 5).

Wellman, et al.*’s* (2001) meta-analysis did examine age differences for ToM tasks within narrow age bands, such that children younger than 4 years performed below chance, children 4 years and above performed significantly above chance. The meta-analysis found no impact of task type – i.e., unexpected identity or change of location - in preschool-aged children. It did, however, find differences for task difficulty in relation to task variations, as mentioned earlier (i.e., tasks involving deception or trickery, child involvement, salience of the protagonist‟s mental state or salience of the contrasting real-world state). More conclusive evidence, in the form of controlled experiments, designed to test for difference in task difficulty, is required to settle this debate.

Further support for the proposal that variations in FB tasks can lead to variations in task difficulty is provided by findings here. Although the interaction with order of presentation was not statistically significant ($p < .05$), this reflected the small sample sizes for the order types, and examination of effect sizes lends some support for staggered test item difficulty. The pattern of order effects resembled training effects, where easier items, presented first, result in better performance on subsequent more difficult items. Furthermore, the reverse (i.e., more difficult items presented first) appears to have little to no impact on subsequent easier items.

In summary, the findings of this study illustrate the need for standardised methods of testing ToM. Difficulty of false-belief tests, which target children‟s ability to represent the false beliefs of others, can vary as a result of relatively minor as well as substantial task variations. In the broader scheme of things, meta-analyses have provided evidence to demonstrate that FB tests are robust to task manipulations in measuring the age at which children perform above chance on these tests. However, individual studies utilising FB tests, often with small samples, can find differences in children‟s competencies, which could be due to task artefacts. This can cause confusion in the literature, redundant debate, and reduce confidence in test use.

CHAPTER 7 - General Discussion

In this chapter the contribution of key findings, presented in this thesis, to the understanding of ToM and its development will be discussed. Firstly, the way in which key findings from the thesis contribute to understanding the development of ToM in relation to language and reasoning will be addressed. Particular attention will be given to the change in factors most related to success on the FB task from the ages of 4 years to 6 years. Secondly, in this chapter a comparison will be made between the pattern of abilities in normally developing 4-year-olds and 6-year-olds and the pattern of abilities observed in children with autism who participated in the pilot study (presented in Chapter 3), in order to assess how typically developing ToM compares to developing ToM in autism. In particular, whether ToM success in autism resembles emerging ToM in typically developing 4-year-olds or whether ToM in autism resembles more developed ToM found in typically developing 6-year-olds will be considered. Thirdly, this chapter expands discussion of the results from the final study; that is, the effect of two differing types of test presentation methods on corresponding ToM scores in 4-year-old children. Limitations and strengths of the current research will be considered alongside the reflection on findings. Finally, consideration will be given to how the findings presented in this thesis fit with and contribute to the debate of theories of ToM; with a focus on how early development differs in typically developing children and children with autism. While the focus is ToM, discussion will include conjecture about the related development of language and cognition. It will be argued that children"s early experiences are fundamental to their development, and that promoting these experiences leads to better outcomes for children with autism. Finally, future directions for ToM research and recommendations for future research will be considered.

Early and Later ToM: Differences and Similarities

The findings presented in Chapter 4 of this thesis indicated that language and reasoning are important for success on ToM tasks, but that the degree of importance changes between the ages of 4 and 6 years. In particular, 4-year-old ToM success was most related to language ability, whereas 6-year-old ToM was related most strongly to reasoning ability. This was a new finding because this type of comparison has not been made before. It is imperative to note that the current findings were associated with a ToM task battery that seemed unusually difficult, even for 6-year-olds in the study. A combination of factors is likely to have contributed to excessive task difficulty; factors which will be examined after the current findings are discussed in relation to the literature.

The current findings add to the understanding of the relationship between ToM ability and language. While the importance of language for ToM has been well established in the ToM literature, the literature has largely not explored the way in which the contribution of language changes over time (Milligan, Astington, & Dack, 2007). Nevertheless, a review of studies that did control for age found a reduced relationship between ToM and language once age was controlled (Milligan, et al., 2007). In line with these findings the research presented in this thesis, and specifically in Chapter 4, has demonstrated that general language ability can mediate performance on false-belief tasks in 4-year-olds. By the age of 6 years, however, language differences between children did not mediate ToM test performance. A potential explanation for the change in the role of language follows.

As discussed in Chapter 2, the proposed role of language takes two general forms: a linguistic competence approach and an early environmental experiences approach. The emphasis in the linguistic approach varies, but generally it links ToM competence with some aspect of language ability that is thought to be necessary for success on ToM tasks. This approach has demonstrated that a range of linguistic abilities are associated with ToM (Milligan, et al., 2007). From the perspective of linguistic competence, the strength of the relationship between early ToM ability and language ability should lessen as children become more linguistically able. For instance, not all 4-year-olds will have developed the necessary language skills to demonstrate an understanding of other people"s minds, especially as it is related to success on false-belief tasks. For that reason, children with better language ability -

that is, language competence as it relates to understanding and responding to false-belief tests - perform better on these tests. Thus, by 6-years-old most children have developed the language skills necessary for false-belief task success, which began to emerge around 4 years of age. It is, therefore, logical that at 6 years of age language ability does not mediate children‟s ToM success as differentially as in 4-year-olds. Hence, in older children language should not correlate as strongly with ToM, despite still being important for ToM.

While a minimum level of language competence undoubtedly plays a role in ToM success at any age, some other ability might better discriminate older children who score higher on ToM task batteries. This other ability might be an actual understanding of others‟ minds or how much the child engages in thinking about others‟ minds. Alternatively, or in addition, a cognitive factor or combination of cognitive factors that develop later (i.e., after ToM begins to emerge at 4-years-old) might contribute to false-belief task success (e.g., attention, memory, reasoning, etc). The findings from the current study have suggested that reasoning ability, specifically as it applies to reasoning about counterfactual situations, is a reasonable mediator of success on ToM tasks in 6-year-old children. While subtractive reasoning scores correlated moderately with ToM scores, factor analysis found that scores for ToM and subtractive reasoning tasks loaded strongly and exclusively on a single factor, which was best thought of as a counterfactual reasoning factor. This finding has extended research that has demonstrated that the ability to reason about counterfactual information is necessary but not sufficient for ToM success (Peterson & Bowler, 2000; Riggs, Peterson, Robinson, & Mitchell, 1998). The current factor analysis established that, in 6-year-olds, the skills required for subtractive reasoning and false-belief reasoning were differentiated from the set of abilities encompassed by language ability, intelligence, and cognitive ability.

Hence, the data from the current study, presented in Chapter 4, demonstrated that the relationship between counterfactual reasoning tasks and ToM tasks was not the result of common language, developmental cognitive ability, or intellectual ability mediating performance on both types of tasks. The factor analysis, in Chapter 4, demonstrated that abilities begin to become more distinct and differentiated with age; a finding that has long been recognised in developmental intelligence research (Anderson, 1992). Compared to 6-year-old children, 4-years-olds abilities were much less differentiated; ToM was strongly related to language, cognitive ability, intellectual ability and subtractive reasoning. In summary, the research findings demonstrated that between the ages of 4 and 6-years children‟s competencies across a range of abilities contributed differently to success on ToM tasks, and that these abilities were not limited to an understanding of false-belief.

While it has been demonstrated that a child‟s level of counterfactual reasoning ability and communicative competence can mediate performance on tasks designed to assess his/her ToM, the way in which ToM develops is also related to the child‟s early communicative environment (de Rosnay & Hughes, 2006). For example, longitudinal research has demonstrated the role of early language ability for later ToM (Astington & Jenkins, 1999; Tager-Flusberg & Joseph, 2005; Taumoepeau & Ruffman, 2008). Furthermore, the environmental approach has demonstrated that a richer tapestry of early environmental experiences can compensate for poorer language ability, bolstering ToM success (Jenkins & Astington, 1996). That is, children were able to pass ToM tasks with lesser language ability if they came from a larger family where, presumably, they were exposed to more sibling interaction and thereby more interactions to provide opportunities to master a ToM (Perner, Ruffman, & Leekam, 1994). Although the studies comprising this thesis did not set out to measure the effect of early communicative environment, a comparison of the pattern of abilities related to ToM in autism to those in typically developing children can contribute to an understanding of the development of ToM in relation to environment. Specifically, comparing abilities related to ToM success in children with autism and typically developing 4 and 6-year-old children may highlight differences in the ways in which false-belief success is achieved in children with autism. This comparison follows.

ToM in Autism: How different is it?

As discussed in Chapters 1 and 2, researchers have argued that ToM in autism is not like ToM in typically developing children. Nevertheless, research has found similarities in the relationships between ToM and other abilities in children with autism and typically developing children. For example, researchers have proposed that, in autism, more advanced language ability is required before ToM success becomes possible (Happé, 1995). Thus, it has been proposed that when ToM does develop in autism the cognitive processes used to succeed on ToM tasks are different (e.g., Frith, Morton, & Leslie, 1991); or that children with autism have poor early social orienting skills (Charman, et al., 1997; Dawson, et al., 2004; Leekam, Lopez, & Moore, 2000); and that processing social information during social interaction is impaired in people with autism (e.g., Reddy & Morris, 2004; Williams, 2004). It should be noted that, while ToM deficits and differences are not unique to autism (e.g., schizophrenia; Brüne, 2005; and deafness; de Villiers, Astington, & Baird, 2005; Peterson, 2004), ToM in other atypical populations will not be discussed here, because this thesis did not examine ToM in these populations. Aside from differences, the literature also reports similar, but not identical, relationships between ToM and other abilities for children with autism and typically developing children; for example, the relationship between ToM and executive functions (Hughes & Ensor, 2005; Moses & Sabbagh, 2007; Müller, Zelazo, & Imrisek, 2005; Ozonoff, Pennington, & Rogers, 1991; Russell, Hala, & Hill, 2003); the importance of language for ToM (Happé, 1995; Milligan, et al., 2007); the role of reasoning in ToM success (Peterson & Bowler, 2000; Riggs, et al., 1998); and the relationship between social competence and ToM ability (Astington, 2003; Dissanayake & Macintosh, 2003; Tager Flusberg, 2003). A discussion follows of how the current findings, presented in Chapters 3 and 4 of this thesis, contribute to understanding these key differences and similarities in ToM.

It has been shown that the pilot data for typically developing children, presented in Chapter 3, fit reasonably well with the data generated from the study presented in Chapter 4. Both studies found that language ability was most related to ToM. The study presented in

Chapter 4 also found that, after controlling for age, subtractive reasoning ability was correlated with ToM when data from 4 and 6-year-olds was combined (note that the pilot study did not include a subtractive reasoning measure). Additionally, the study in Chapter 4 was specifically planned to examine the relationship between ToM and developmental abilities within age cohorts. As previously discussed, this analysis showed that language was most strongly related to ToM at 4 years of age, whereas only subtractive reasoning was related to ToM at 6 years of age. Moreover, factor analysis demonstrated that 4-year-old ToM, language competence, intelligence, and cognitive developmental ability shared an underlying factor, proposed to broadly influence development. In comparison, however, the range of abilities measured in 6-year-olds was beginning to become differentiated; a broad developmental factor continued to exert an effect on language, cognitive ability, and intelligence, but ToM and subtractive reasoning did not load on this same factor and were loaded together on another factor. In light of the similarities between the data for typically developing children from these two studies, an evaluation of the data from typically developing children taken from Chapter 4, differentiated by age groups, are here compared to the pilot data from the children with autism.

Children with autism who passed the false-belief tasks in the pilot study had on average significantly better verbal ability (as measured by the BDI Communication Domain) than those who did not. Differences in cognitive ability were also evident and, despite lack of power, these approached significance ($p = .07$). This outcome compares best with the pattern of results for normally developing 4-year-old children from the Chapter 4 study. In contrast, the pattern of data from the normally developing 6-year-old children do not fit with the data from the children with autism. A plausible suggestion, then, is that ToM ability in the sample of children with autism most resembled the performance of normally developing children when early ToM is emerging. Of course, the sample size of the pilot study, together with the population sampled (half the children had received early intensive one-on-one intervention), limits generalisability of these findings. Nevertheless, this is an intriguing possibility, which warrants further examination. Although ToM in autism has been thought to be achieved differently to ToM in typical development, the current findings suggest that this is not accurate or specific. At the very least, it is probable that, with the necessary input, children with autism‟s ToM develops comparably to ToM in typically developing children, albeit later.

It should be noted that the pilot study and subsequent larger study differed in a number of ways. Firstly, the pilot study included data from a sample of children ranging in age from 3 – 7 years, whereas the subsequent study analysed data from children aged either 4 or 6 years. Secondly, the pilot study used a single ToM task, whereas the subsequent study used a battery comprising four ToM tasks. These differences do not appear to have contributed to differences in findings; although the difference in the number of ToM tasks did result in different methods of data analysis. If anything, the data from the second study is likely to be more robust. Furthermore, the pilot study included a social developmental measure that could not be included in the subsequent study. This measure involved interviews with teacher or parents, which was not practicable for teachers in the second study, who had a large number of children in their classes involved in the research and therefore did not have the time to complete this aspect of the research. Future researchers, with more adequate resources, should consider including measures of social competence to assess how these relate to ToM in different populations.

Indeed, it would be interesting to compare early and later ToM in children with autism to early and later ToM in typically developing children. This type of comparison could shed light on the respective routes of ToM development in these divergent cohorts. While longitudinal research for each population has been reported in the literature (Astington & Jenkins, 1999; Tager-Flusberg & Joseph, 2005), these studies have not examined the same factors, therefore limiting comparison.

Although the findings of the current studies have not diverged extensively from the literature, the perspective of the current research differs from that generally found in the literature, in that is does not suppose that the nature of ToM is radically different in autism.

Instead, it is proposed here that developmental differences in fundamental cognitive processing, as opposed to a specifically deficient ToM mechanism, may play a more primary role in understanding autism than has previously been widely accepted. That said, it is imperative to acknowledge that this is not a new suggestion and others have made similar claims (Bjórne, Johansson, & Balkenius, 2006; Bogdashina, 2005). Nevertheless, ToM research in autism that has focussed on a specific ToM deficiency is highly prevalent and convincing. Perhaps the reason why ToM has been highly emphasised in autism, more so than in other conditions where it has also been found to be different or deficient, is that autism is mostly thought of as a social, communicative disorder (DSM-IV; American Psychiatric Association, 1994; Gotham, Risi, Pickles, & Lord, 2007; Lord, Rutter, DiLavore, & Risi, 1999; Lord, Rutter, & Le Couteur, 1994). It is certainly the case that high functioning individuals with autism appear especially socially inept and find social interaction very difficult (Williams, 2004). However, it is much more difficult to think of autism in young children solely in these terms.

Autism is a developmental disorder and young children with autism display a much broader range of symptoms; symptoms that cannot be explained by a lack of ToM alone. These symptoms include, for example, sensorimotor disturbance; attention deficits; sleep problems; repetitive behaviours; and difficulty generalising knowledge (Bjórne, et al., 2006). The underlying cause of these symptoms - for arguments" sake, a wide spread sensory processing disorder - might also disturb normal development of ToM. For instance, ToM in autism has been linked to symptom severity in two of the diagnostic categories (social and communicative), but not to daily living skills or functional communication (assessed by the Vineland Adaptive behaviour scales) (Tager Flusberg, 2003). The study did not investigate the relationship between ToM and the third diagnostic category, repetitive behaviour and, therefore, it is not possible here to comment on these. Repetitive behaviour, however, is now the most relevant domain and such analysis could help establish whether ToM can account for all symptoms of autism, rather than those that share a social/language component. Some evidence that supports the proposition that ToM is not central to all the symptoms of autism comes from research that has examined the relationship between early social abilities and the core symptoms of autism (Delinicolas & Young, 2007). This research found that deficits in joint attention (a vital early social ability important for later development of social skills) were most strongly related to core social and communicative deficits, more so than the non-social core symptoms (i.e., repetitive behaviours and insistence on routines).

It is of course likely, as suggested in the literature, that a number of physical dysfunctions cause the range of symptoms of autism (Frith & Happé, 1994; Jarrold, Butler, Cottington, & Jimenez, 2000; Morgan, Maybery, & Durkin, 2003; Pellicano, Maybery, Durkin, & Maley, 2006), and differentiating between these accounts will require much more research. Future research should consider symptoms of autism that are shared with non-autistic disorders; specifically in relation to the likely causes of those symptoms in non-autistic disorders. For example, the likely cause of a ToM deficit in children born deaf to hearing parents is the lack of necessary communicative experiences required for the development of ToM. Given that children with autism also have limited early communicative experiences - most likely stemming from a processing deficit that limits successful interaction with the environment - the most logical conclusion to draw would be that a lack of successful early interaction experiences has detrimental effects for later ToM. Evidence from neurological research should be considered in the formation of any theories of ToM or autism. While this does currently happen, even cursory examination of this literature confirms that there is very little agreement as yet about the neurological bases that ultimately must underpin poor ToM. Nonetheless, it is useful here to examine more closely the way in which early communicative experiences differ in children with autism when compared with those of typically developing children. It is proposed that this is useful for examining later ToM differences and similarities in these cohorts.

However, before proceeding to a discussion of the ways in which early communicative experiences differ between children with autism and typically developing children, this chapter will first discuss limitations and strengths of the current study, primarily in relation to the measures used across studies.

Reliability of ToM Measures

This thesis set out to examine ways in which ToM was related to development in children with autism and in typically developing children. Although data were collected for typically developing 4-year-olds and for typically developing 6-year-olds, the planned data collection with children with autism was initially postponed, in order to investigate the reliability of the ToM measures that were to be used in the first two age groups. The decision to postpone planned data collection was made as a result of seemingly unusual ToM task difficulty. Thus, instead of the study planned initially, an alternative study was designed to assess a possible cause of difficulty with the ToM test battery. The findings from this study were presented in Chapter 6. These findings are reviewed here and discussed in relation to the research reviewed in earlier chapters of this thesis.

In the first instance, a review of the ToM task battery suggested that task difficulty may have been related to the way in which the tasks were presented. Although there have been variations in the way false-belief tasks have been presented, much research has utilised dolls and props to illustrate the story to children (Wellman, Cross, & Watson, 2001). In the study presented in Chapter 4, stories were read to children and illustrated with a cartoon strip. Although previous research had reported that presentation method does not increase ToM task difficulty (Wellman, et al., 2001), it is possible that something about the cartoon strip was particularly confusing or challenging. A follow up study, the findings from which have been presented in Chapter 6, demonstrated that the cartoon strip tasks were more challenging for 4-year-olds than similar tasks presented with dolls and props. The difference between Wellman et al.‟s findings and the current findings are likely to stem from differences in research methods. Wellman‟s results demonstrated that, on average, and over a large number of studies, differences in ToM scores in relation to task presentations (i.e., using either pictures or props) were negligible. On the other hand, current findings have clearly demonstrated that presentation method can have an effect on task difficulty that may be important, especially for young children. The difference in difficulty between task presentation types was not negligible and certainly sufficient to propose that this was a difference that can lead to confusion about findings and cast doubt on the quality of testing procedures, sample selection, and so forth.

Further review of the data from the study presented in Chapter 4 revealed that perhaps something about the tasks in which people changed location was particularly difficult for children age 4 and 6-years. These tasks were on average failed more often than tasks in which objects changed location and tasks which involved no change of location. There are two possible explanations for this difficulty. Firstly, children in the age groups assessed might have had some specific difficulty reasoning about people changing location because of a characteristic attached to people but not to objects used in the stories (e.g., people move themselves, they are independent agents; they make noise and might, therefore, be found more readily; they are larger and, therefore, more easily seen; they have beliefs or desires that caused them to change locations). Any one of these characteristics or a combination of these characteristics might make reasoning about where a person will look, upon his/her return, for another person more challenging. The second possibility is that the cartoon strips involving people changing location were particularly confusing (a possibility suggested later by an anonymous reviewer). The data from the study presented in Chapter 6 can be used to differentiate between these two accounts.

An examination of the pattern of difficulty across the two types of presentation method (i.e., 2D cartoon strips and pictures vs. 3D dolls and props) revealed that the first account was more plausible. Children aged 4 had some specific difficulty reasoning about people changing location. While, on average, the 3D presentation type was easier across all three types of ToM tasks (i.e. object changes location, or person changes location, or no change of location), it was only significantly easier for the *object change of location* and *no change of* *location* tasks. This finding does not lend support to the possibility that the cartoon strips were particularly confusing, because the same difficulty was evident when using dolls which moved from one location to the next (the same procedure used for standard ToM tasks). This analysis did not form part of the original analysis conducted for the manuscript presented in Chapter 6, and is instead presented here in order to clarify the nature of the task difficulty (see Table 1).

Table 1.

Comparison of 2D and 3D Presentation Types Separated by Type of ToM Task

	2D score	3D score	CI_{95} for mean difference between 2D and 3D	
Task type	M (SD)	M (SD)	Upper	Lower
Person changes location	.60 (.81)	.83 (.87)	-.55	.087
Object changes location	.77 (.86)	1.10 (.92)	-.60	-.068
No change of location	1.57 (.73)	1.80 (.48)	-.47	.00

Note: N = 30; possible range of scores for each type of ToM tasks was 0-2

It should be noted that the tasks in which people changed location were not second-order ToM tasks. Second-order tasks are thought to measure later ToM, and, hence are termed advanced ToM tasks (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Brent, Rios, Happé, & Charman, 2004; Kaland, Callesen, Moller-Nielsen, Mortensen, & Smith, 2008). Second-order tasks are believed to assess a person‟s ability to infer the thoughts of another person about the thoughts of a third person. Aside from the added complication of thinking about two people‟s thoughts, these tasks are certainly more cognitively complex. They involve more complex stories, involving a series of changes. This alone must place increased demands on working memory. For instance, research found that an increased number of events in a causal chain impacted negatively on children‟s counterfactual reasoning (i.e., reasoning about possible outcomes considering that a specified event had not taken place) (German & Nichols, 2003). The tasks used in the current author‟s study, however, did not involve more events than standard ToM tasks. Nor did these tasks involve second-order reasoning. For that reason, and as has been proposed above, the difficulty is most likely to stem from a characteristic of people that children find difficult to process at this age. It is perhaps the same characteristic that increased difficulty of some tasks for children in Russell et al.‟s modified windows task, which has been discussed in Chapter 2 of this thesis (Russell, et al., 2003). Specifically, having a competitor present increased task difficulty for children with autism, even when children were not required to deceive the competitor.

Children‟s difficulty with ToM tasks in which people change location is, therefore, likely to result from a characteristic unique to people and absent from objects. The current data set cannot be used to examine what this might be, but future research should take this into account when designing ToM tasks. For example, an exploration of the nature children‟s difficulty with tasks involving people changing location might, in the first instance, ask children to qualify their answers to the ToM test question for both *people change of location* and *object change of location* tasks. Qualitative data collected may reveal that children have a specific tendency to reason based on a set of characteristics inherent to people but not objects.

In addition to the finding that 2D tasks were more difficult for 4-year-olds, the data also suggested that the order of presentation might affect overall ToM success. Although the analyses found no statistically significant differences in ToM success as a result of order of presentation, the study was underpowered for this comparison and a clear pattern of differences was evident, whereby administration of easier tasks first resulted in overall higher ToM scores. Order effects resembled practice effects, whereby early presentation of easier tasks had a positive impact on performance for later more difficult tasks, but not the reverse.

The current analysis of presentation methods is the first direct experimental comparison. This analysis adds to the understanding of the way in which difficulty of tasks, believed to measure ToM, can be affected by small, seemingly negligible changes. While previous research has found that on the whole this poses no problems for theories about the nature of ToM (Wellman, et al., 2001), differences between presentation methods clearly do impact task difficulty. A benefit of the current analyses is that the study‟s repeated measures design, and task equivalence (for 2D and 3D tasks), allowed for more direct comparison by eliminating potential confounding factors. Thus, the current analysis highlighted the importance of developing a standardised measure of ToM (examples of which have been discussed in Chapter 5 of this thesis).

In short, using tasks that vary, even in seemingly insignificant ways, can produce confusing results or, more radically, produce results counter to what is expected. When this happens other factors might be thought to contribute to findings; such as, confusing test materials, poor sampling, poor test administration procedures, and so forth, all of which bring into question the integrity of the research methods employed. This, in turn, can limit publication of potentially important findings. In summary, it is important to ensure that tasks used across studies are comparable in order to limit potential confounding factors, and allow for less complicated comparison of research findings.

The Development of ToM: The Role of Language, Cognitive Ability, and the Child's Social Environment

Earlier in this chapter the contribution to the literature of the findings constituting the basis of this thesis were discussed. The research methods used in the current body of studies have limited the discussion of contributions to this literature. Specifically, the current findings were in relation to language and cognitive competence as these related to ToM competence. Therefore, the findings cannot speak to longitudinal relations between these variables. In spite of this limitation, the body of findings discussed, heretofore, can be thought to support the theory that a disturbed early social environment – resulting from impaired processing ability that limits successful interaction with the environment - results in children with autism‟s ToM deficit and delay. Moreover, this is the theory given most credence in this thesis. For this reason it is fitting to discuss how the early social environments of children with autism vary from those of typically developing children.

Astington and Baird (Astington & Baird, 2005) highlighted the complex relationship between language and ToM. The authors proposed that a child‟s language competence at any given time was likely to influence the type of environmental experiences that the child was exposed to, which in turn influenced how s/he developed both ToM and language skills. Discussion of findings from both the literature and this thesis, that support this claim, is presented here. Longitudinal research reported that the relationship between ToM and language was such that language ability promoted ToM, but that ToM did not promote language ability (Astington & Jenkins, 1999). However, Astington and Jenkins‟ research did not measure the influence, on language, of abilities thought to precede ToM competence; a set of abilities which, Nelson argued, are important for psychological growth (Nelson, 2005). It is very likely that these abilities (e.g., joint attention, social orienting) are fundamental to the development of both language and ToM. Additionally, the role of conversation (i.e., communicative environment) has been demonstrated as important for development of ToM (Astington & Baird, 2005; Brown, Donelan-McCall, & Dunn, 1996; de Rosnay & Hughes, 2006; Dunn, Brophy, Astington, & Baird, 2005; Peterson & Siegal, 1995). In line with this range of findings, the results comprising the body of outcomes from the current thesis have indicated that language is important for early ToM. Language competence contributed to performance on a ToM task battery in 4-year-olds, when language ability was less mature. Hence, the factor analysis, reported in Chapter 4, might be thought to indicate that early communicative environment influences both the development of language ability and ToM, because both of these abilities correlated substantially with a developmental factor. In fact the developmental factor likely reflects the child‟s early environment. Similarly, the factor analysis of 6-year-old abilities demonstrated that alternative competencies contributed to a more mature ToM. So, while language is important for the development of ToM, the extent to which language correlates with ToM, at a given time, depends largely on whether a minimum level of language has, on average, been achieved in that age cohort. The present findings, taken together with the findings from the literature discussed here, indicate that early communicative environment is most important for the development of ToM. Put another way, *competence* influences the type of experiences a child has at any given time and those experiences (*environment*) influence how the child continues to develop.

Tager-Flusberg and Jospeh (2005) proposed a different route for ToM development in autism. The route they proposed in typically developing children is one in which a "social-cognitive component of theory of mind builds on the earlier emerging perceptual component" (p. 311). That is to say, children"s ability to reason about others" thoughts relies on an earlier ability to perceive social cues. In contrast, children with autism are thought to have "a fundamental deficit to read and use the social-perceptual information available from faces, voices, or body gestures, that is, the social-perceptual component of theory of mind" (p. 311). Tager-Flusberg and Joseph proposed that children with autism who eventually succeed on ToM tasks bypass the typical route by compensating with language abilities. That is, children with autism use more mature language skills to prop up reasoning about social phenomena. Conversely, the data from the studies making up this thesis have demonstrated to some degree that development of ToM in autism was not altogether different to development of ToM in neuro-typical children. Namely, the skills related to early ToM in typically developing 4-year-old children were comparable to those that were found to be related to ToM in children with autism.

Nevertheless, it is clear that children with autism have very different early abilities and as a result vastly different experiences than do typically developing children. It is argued here that these experiences shape later ToM and, that if we can better establish ways to promote these experiences in autism, later ToM should not be vastly different from that of typically developing children. As previously discussed, the data from the pilot study presented in Chapter 3 were drawn from a sample of children with autism, half of whom had received early intensive one-on-one intervention. Intensive one-on-one intervention might explain the apparent similarity between the ToM of children with autism and the typically developing children. Specifically, the one-on-one nature of early intervention might have provided children with autism with enough communicative exchanges to contribute positively to ToM. Twin research with typically developing children supports the conclusion that this type of experience is important for ToM. Hughes and colleagues (Hughes, Astington, & Baird, 2005) have reported that typically developing children‟s early skills (e.g., joint attention, imitation, affective contact) are largely innate and universal and that later social environment plays a larger role in subsequent ToM because such skills are intact. Children‟s experiences outside the home, once attending school or preschool, were thought to add differentially to ToM in twins, who were likely to have similar experiences at home. An alternative explanation is that the one-on-one intervention simply increased children‟s verbal ability, and in this way propped up deficient ToM ability. However, children in the current sample succeeded on the ToM task at a much earlier verbal mental age than that reported necessary in the literature (Happé, 1995). Therefore, it is unlikely that verbal ability alone could explain better ToM in this sample.

While a number of researchers are not likely to agree with the following, preferring a more innate or modular account of childhood development, it is argued here that researchers need to work toward a cohesive theory that encompasses the role of the environment. This may be in the form of an account of how damage to neurological systems impacts on children‟s interaction with their environment. Chiefly, this is necessary because the literature has clearly demonstrated that both linguistic devices and conceptual understanding result from a child‟s environmental influences, in this case the child‟s early communicative environment. This reviewer‟s perhaps simplistic but nonetheless common sense explanation is derived from a synthesis of the many varied research findings and from first-hand experience in working with children with autism and typically developing children. While it is possible that an inborn preference for social stimuli reflects an underlying mental mechanism, it is also possible that necessity dictates the types of early interactions children are both exposed to and initiate. Using the example of *want* and *think*, children are typically exposed to, and therefore become proficient at *want* well before they are exposed to, and therefore become proficient at *think*. As any parent will attest, from birth children are self-centred desire-based beings and parents, while also initiating other interactions, largely respond to the child‟s needs. These needs may take a physical form (e.g., hunger or cold) or a psychological form (e.g., affection or attention). For this reason, early utterances and a large proportion of early communication interactions revolve around fulfilling the child‟s needs or wants. Corroborating scientific evidence comes from research that has empirically demonstrated what most parents had long before experienced first-hand, that is that children understand *want* long before they understand *think* (Wellman & Liu, 2004). It is also likely that the differing nature of the referent objects of *want* and *think* contribute to the ease with which these concepts are learned. As pointed out by de Villiers, et al. (2005) *want* usually has a concrete reference (an object), whereas the referent of *think* is ordinarily more abstract.

It might, however, be argued that not all communication is needs based or environmentally driven. Children with autism demonstrate pure needs-based communication, in that they often exhibit functional communicative attempts (e.g., expressing desire for an object), although they rarely exhibit communication attempts to share interest (Charman, et al., 1997). Conversely, young typically developing children demonstrate a range of communicative behaviours, such as those that fall under the term joint-attention; ranging from gaze following to orient to another person‟s focal point through to pointing to orient another person to something that the child finds interesting (Charman, et al., 1997). While these behaviours may not appear to relate to any physical need, it is argued here that they correspond to a biological need; that is, an innate imperative to learn. Learning about the environment and making sense of the world are certainly important for successful functioning. There are two prime candidates for the way in which children do this; firstly, interacting with „knowing‟ adults and, secondly, a set of behaviours that are commonly referred to as *play*.

Evidence in support of this experiential account of children‟s acquisition of mental state understanding comes from children‟s early pretence behaviours (Lillard, 1993). These behaviours form part of children‟s *play* repertoire. Pretence can be thought of as one way in which children interact with and learn about their environments. Children are exposed to pretence in day-to-day *play* much earlier than they hear about and initiate talk about thinking (Lillard, 1993); consequently, children master pretence well before mastering the concept of thought, as measured by standard tests of ToM. As Lillard explains, children demonstrate a higher level of functioning in pretend play than they do in non-pretend situations. Indeed, the current author has observed young children engaged in pretend play and witnessed how children use pretend play to rehearse aspects of their early environments. For example, a 2-year-old girl was observed practicing talking into a telephone to her grandmother in pretend play before being able to competently talk into a telephone to her grandmother in non-pretend situations. It is most probably the case that the young girl was less cognitively taxed by the pretend situation; the pretend situation advanced at a pace governed by the child‟s ability and the child only used language that she comprehended. Observation of the same child actually talking on the telephone, to her grandmother, a few months later highlighted how the earlier „practice‟ enabled later real interactions.

There is nothing ground-breaking about the idea that exposure to a concept leads to learning about that concept; and that earlier exposure results in earlier understanding. From this perspective, the lack of *play* behaviours in autism can be thought to lead to impoverished learning about the environment. *Play* behaviours vary greatly over the lifespan. Earliest *play* behaviours lead to learning about the child‟s own body (e.g., cooing noises made by infants); learning then extends to the immediate, and then eventually to the broader environment. For example, some early play teaches children about cause and effect (e.g., hitting at a mobile); engaging with parents and peers in later play can teach children language skill (e.g., reading a story or shared manipulation of a toy); physical play helps children develop gross-motor and fine-motor skills; and so forth. The repetitive nature of children‟s play might be thought as rehearsal, leading to improvement in skills.

Conversely, children with autism do not engage in play, and particularly pretend play in the same way (Charman, et al., 1997; Gotham, et al., 2007; Lord, et al., 1994). In part this may be because children with autism do not imitate their environments (Charman, et al., 1997); a lack of imitation might be the result of impoverished observation of the environment or the result of a missing innate drive to imitate the environment. Poor observation of the environment leading to impoverished imitation is a more likely explanation, if a sensory processing deficit is considered responsible for some of the symptoms of autism. Children with autism do, however, exhibit some of the fundamental factors associated with typically developing children‟s *play*. That is, children with autism exhibit repetition and imitation, albeit repetition of maladaptive behaviours and limited imitation such as echolalia (DSM-IV; American Psychiatric Association, 1994; Gotham, et al., 2007; Lord, et al., 1994). These limited and different manifestations of repetition and imitation might be considered to reflect the same innate drive for learning that is present in typically developing children. It is conceivable that differences in the manifestations of these drives could be ascribed to sensory processing deficits and differences, in autism; differences that prevent children from focusing on and learning about relevant stimuli. An inability to process information in a coherent whole was proposed by Frith (1989) to explain some of the symptoms of autism. The central coherence theory, as it is termed, proposes that people with autism process information in a segmented, piecemeal way rather than as a global whole. This account fits with a sensory processing deficit in autism, if it is considered that a tendency to process information in this way may be the brain‟s way of dealing with limited processing capacity. Recent neurological research supports the proposition that normal sensory processing is disrupted in autism (Medrihan, et al., In print).

In summary, a possible account for children‟s development of skills and abilities, as these are related to human environments, has been presented. This account emphasises the fundamental role of child-initiated *play* behaviours in both typical development and autism. Differences in the manifestations of these behaviours are proposed to account for differences in later skills and abilities. While this account may seem overly simplified, it is in fact more cohesive because it captures the way in which neurological systems interact with the environment. The account emphasises the importance of both environmental experiences and neurological systems. Thus, the strength of the account is that is explains a complex set of behaviours in relation to a fundamental feature of the human brain (i.e., information processing).

Future Directions

The research, summarised in the current chapter, has demonstrated the need for a better understanding of how children‟s early abilities relate to ToM differentially across clinical and non-clinical cohorts. A better understanding of these relations may be achieved by using more targeted measures, rather than general measures, along with ToM measures that are both reliable and demonstrate good construct validity.

Reliability of measures should be assessed in a number of ways. Firstly, short-term test-retest reliability must be assessed to insure children can perform consistently on the measure across weeks and months. Test-retest reliability research must take into account the age of population tested. A test battery might unduly tax children of a particular age group, so test-retest reliability should be established for all age groups with which the measure is intended for use. For example, limited attention spans in 3-year-olds might result in poor test-retest reliability within this age group, but 6-year-olds might have no such difficulties. ToM research has focused on young children who are still learning about concepts of mind, therefore when creating tests researchers must consider the limitations of young children. Limitations include, short attention spans, fluctuating moods, easily distracted, require breaks to maintain focus, not always cooperative, limited language comprehension and expression, etc. It is highly likely that the most appropriate tests of ToM, which are more sensitive to the cognitive and behavioural limitations of very young children, will be based on both observation of the child and experimental assessment of the child. In this way, it is likely that tests will have better construct validity and be more reliable.

Researchers generally spend brief amounts of time with children whom they assess and it is difficult in this time to obtain an accurate picture of the child"s abilities, behaviour, and so forth. Children behave differently in different situations. How they behave is related to how comfortable and safe they feel in the immediate environment. Some children become shy and withdraw, while others test the boundaries of the new environment. For this and many other reasons, it is vital that any true assessment of children"s understanding of other people is made in an environment where the child feels at ease to communicate his/her understanding. For example, a ToM assessment might include structured play activities that the research initiates within the child"s usual environment. There are some ethical considerations here related to the involvement of children in research that may not be the target of the assessment but who would still interact with the researcher within a research setting. These might be overcome by creating play opportunities within the usual physical environment for a small group of children whose parents have given consent. It is likely that these measures would have better construct validity, but it would be necessary to account for interrater reliability. Observational measures create more opportunity for subjective judgements, and as such these types of measures need to be carefully assessed for "assessor" related limitations.

ToM research with older children and adults might not need to be so heavily reliant on observation. The performance of older children and adults on a measure is less likely to fluctuate drastically, if the measure is constructed adequately. It is still necessary, however, to establish the construct validity of any such measure. Establishing construct validity might involve observing real-world interactions requiring use of ToM (e.g. negotiating with a peer, settling a disagreement, sharing an obscure joke, etc.) and comparing these observed abilities to experimental, laboratory measures of ToM (e.g. an test battery of false-belief type tasks). In order to be considered useful experimental ToM tests should also demonstrate a degree of predictive validity. Although a test battery might be faster and easier to administer, in both research and clinical practice, than an observational ToM measure, observational items might be included for increased construct validity. In other words, the experimental measures should tell us something about how the person actually uses his/her ToM in real interactions. Gaining an understanding of how people use an understanding of other people‟s thought processes should be a high priority for ToM research. An understanding of how performance on ToM test batteries relates to a person‟s actual ToM competence in real-world situations would be very useful for broader application in both clinical practice and research. For instance, a psychologist who has a client who is unable to form and maintain relationships might assess the client‟s ToM with a measure known to reflect actual real-world use of ToM. For arguments sake, the ToM assessment might reveal that the client, while being perfectly capable of reasoning about other‟s thoughts, only takes into account the perspective of another person when s/he is has "nothing to lose" by doing so. Current ToM test batteries are unlikely to be able to assess the use of ToM in this way; instead it could be useful to devise an interview based measure that addresses how a person uses his/her ToM in day to day life.

In brief, the research community must work toward, and subsequently implement, a standardised measure of ToM. Any such measure should be assessed with a large normative sample (certainly much larger than the sample sizes that have been reported in ToM research to date). It is not, however, always possible to recruit such large samples in a single research group, and for that reason it might be necessary for large research centres to work collaboratively. I must note here that it is probably unlikely that research centres will be able to collaborate on such a large scale. A number of issues are likely to limit large scale collaborative work, including but not limited to intellectual property negotiations, differing research agendas, language barriers, and so forth. Nevertheless, large scale collaboration is eventually likely to produce a well-tuned measure. A collaborative approach has the added benefit of bringing together broad ideas and experiences, which can result in a product that is more progressive and better developed than it would have been had it been produced within a single institution that subscribes to a particular theory or methodological approach.

In summary, any new ToM instrument should improve on existing ToM tasks. Firstly, a ToM instrument should assess a broad range of mental state understandings. Secondly, experimental assessments of ToM should have good reliability for use with young children. This might be achieved by having multiple items to assess different constructs, or by including observational test items. Thirdly, ToM instruments for children and adults should be useful in a clinical setting, not just for research. Assessing a person‟s ToM ability must be able to tell a clinician something useful in order to help improve the social life of his/her client. While these requirements of a ToM test are idealistic, it is recognised that there are many factors that can and will limit future research. Therefore, the current author recommends, in the very least, that in the absence of a useful and standardised ToM instrument, researchers should consider the ways in which tests are administered and manipulated, and the ways in which these changes can contribute to differences in subsequent findings.

This pondering has not, however, included a recommendation of how ToM should be thought of and researched in clinical populations. The perspective of the current author has been that while it is possible to reason that lack of a ToM module leads to "no ToM", a more cohesive theory captures the interaction between neurological systems and the environment. This thesis has presented evidence, from both the literature and the studies comprising this thesis, that shows that ToM does develop in autism, and that it can be different and may be achieved differently. However, to a degree ToM in autism does resemble ToM in typically developed children. The process by which ToM is achieved in both populations is most likely through shared communicative experiences in childhood; experiences that can be initiated by other people or the child. An example of child-initiated behaviours, important for development, was discussed in terms of *play*. *Play* behaviours were considered to be manifestations of an innate drive to learn about the environment, a drive initiating from the brain‟s automatic propensity to process and organise stimuli encountered. Future research might examine the ways in which these behaviours can be promoted in children with autism. Ideally, interventions should target sensory processing, but in the absence of technology capable of stimulating or simulating the brains normal function, it is likely that interventions, at least for now, will need to be behavioural or learning based. For instance, some interventions currently work to teach children with autism to imitate a target behaviour. Research might examine how teaching focused imitation impacts on the ability of children with autism to imitate their environments. It is likely that research will find that a key predictor of how well children learn to imitate their environments is directly linked to their ability to generalise the skills they have learnt to novel situations.

Evidenced based interventions for children with autism currently promote learning and functional behaviours. These types of interventions are and need to be intensive. That is, the number of hours spent in interventions is important for learning. By way of comparison, typically developing children spend the majority of their waking hours learning and engaging with the world. It is reasonable, therefore, to propose a great deal of time is required to learn and develop the skills required to successfully interact with the environment. This type of intensity is, consequently, extremely important for children who ordinarily do not interact with the world, perhaps as a result of a sensory processing impairment.

In summary, promoting the development of ToM in children with autism, will undoubtedly involve intensive targeted intervention. Intervention will need to be early to simulate how and when children typically begin to learn about other people. Any intervention will need to include steps for generalising skills and concepts learnt to real-world interactions. Finally, researchers should strive to develop interventions that target the underlying physiological cause of the symptoms of the disorder.

Recommendations for Future Research

The current PhD project was embarked upon with only limited appreciation for the trials and tribulations of conducting research with children. Although the experience gained during the current author‟s honours year emphasised the difficulty of recruiting participants within a short time frame, it was thought that the three years of a PhD would not present the same limitations to data collection that were experienced in the six month time frame for the honours thesis. Despite employing all available and sensible methods, recruiting sufficiently large samples for the studies comprising this PhD took much longer than expected. In hindsight, the current project was probably not suited to a three-year PhD. Upon reflection it is recommended that future research, which aims to assess the issues addressed in this thesis, only be conducted with more time, greater resources (e.g. more research staff), and possibly a larger population base (Adelaide has a population of only 1 million people). Certainly, any attempt to recruit a sufficiently large sample of children with autism – in order to follow up on the findings presented in this thesis with typically developing children – would need to sample from a much larger population.

It should be noted that the ToM test battery used for the present research was perhaps not the best measure of ToM. Although the measure included multiple items to test ToM so that reliability of the test battery would be higher than for single item measures, the test battery should have included items to assess other aspects of mental state understanding. In its present form the measure only assessed false-belief understanding. Although the ToM literature generally concerns itself with measuring false-belief, it is clear that more a comprehensive ToM measure would better detect children‟s developing understanding of mental phenomena. Certainly there have been attempts to create such measures (as discussed in Chapter5), but at the time of designing the current set of studies, descriptions of better measures had not yet been published or published measures were not suitable for the current research. It is recommended that future ToM research uses measures are designed to assess a range of mental state concepts (e.g. belief, desire, false-belief, humour, etc) and that any such measure includes multiple items for each concept.

An overview of the types of difficulties encountered during data collection is also presented here by way of emphasising the types of factors that should be taken into consideration when planning a research project of this kind. Delays to data collection were in part due to the age range targeted in the studies. Children aged 4 and 6 years were recruited through preschools and schools. Although the study had approval from the relevant State Government department, each school principal or preschool director had to approve the study for use within his/her school. No single school had sufficient enrolment (in proportion to response rate expected) to carry out all data collection in a single institution. Hence, the process of enlisting the required number of schools was time consuming and resulted in many months of delays (13 months in total). For the most part schools were approached concurrently rather than in sequence, which limited delays substantially. Concurrent testing across schools was, however, not usually possible. That is, for any given time period, scheduling of visits to assess children was usually limited to one school. This was a result of scheduling clashes. Usually schools had similar timetables and therefore similar times at which children were available for testing. This caused a number of substantial delays. The longest delays were a result of school holiday periods. In addition to not being able to access children during school holidays, generally the week or two directly before or after a holiday period were not convenient times for children to be taken from classrooms for testing. At these times teachers had their own testing to conduct or teachers needed time at the beginning of a school term to settle their classes back into the rhythm of school. Further delays were related to the limited number of hours each day during which children could be taken from their classes. This meant that it was often necessary to wait as much as 1.5 hours to see the next child. As a result it was usually only possible to see two children on a single visit, although when children returned for their second session (where applicable) it was usually possible to see up to three or four children in a single visit. In addition to these delays, inherent to testing children in schools, the usual delays and disruptions to data collection with human populations were also encountered (participants not attending school, participants not cooperative on a given day and assessment postponed, etc.).

In summary, the feasibility of any research project should be carefully considered while planning research. These factors should be given as much consideration as the types of test materials, the rationale for the study, and so forth. Even the best intended and well planned research can essentially be a waste of time and resources if potential limitations are not carefully considered at the outset.

References

American Psychiatric Association. (1994) Diagnostic and Statistical Manual, 4th Edition. (DSM-IV) Washington, DC: American Psychiatric Association.

Amsel, E., Trionfi, G., & Campbell, R. (2005). Reasoning about make-believe and hypothetical suppositions: Towards a theory of belief-contravening reasoning. *Cognitive Development, 20*, 545-575.

Anderson, M. (1992). *Intelligence and development: A cognitive theory.* Oxford, England: Blackwell.

Astington, J. W. (2003). Sometimes Necessary, Never Sufficient: False-Belief Understanding and Social Competence. In B. Repacholi & V. Slaughter (Eds.), *Individual Differences in Theory of Mind Implications for Typical and Atypical Development* (pp. 13-38). New York: Psychology Press.

Astington, J. W., & Baird, J. A. (2005). Introduction: Why Language Matters. In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 3-25). New York, NY US: Oxford University Press.

Astington, J. W., & Baird, J. A. (2005). Representational Development and False-Belief Understanding. In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 163-185). New York, NY US: Oxford University Press.

Astington, J. W., Harris, P. L., & Olson, D. R. (1988). *Developing theories of mind*. New York, NY US: Cambridge University Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*, 1311-1320.

Baron-Cohen, S. (1989a). Are autistic children behaviourists? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders, 19,* 579-600.

Baron-Cohen, S. (1989b). The autistic child‟s theory of mind: a case for a specific developmental delay. *Journal of Child Psychology and Psychiatry, 30,* 285-297.

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

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

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4,* 113-125.

Bjórne, P., Johansson, B., & Balkenius, C. (2006). Efeects of early sensorimotor disorder on contextual learning in autism. *Revue européenne de psychologie appliquée, 56*, 247-252.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25-b31.

Bogdashina, O. (2005). Sensory Issues in Autism are Recognised But... Retrieved 15/03/2006, 2006, from http://www.awares.org/conferences/show_paper.asp?section=000100010001&conferenceCode=000200010014&id=34

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8*, 283-299.

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2008). Theory-of-mind development influences suggestibility and source monitoring. *Developmental Psychology, 44*, 1055-1068.

Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings, and mothers. *Child Development, 67*, 836-849.

Brüne, M. (2005). Theory of Mind in Schizophrenia: A Review of the Literature. *Schizophrenia Bulletin, 31*, 21-42.

Carpendale, J., & Lewis, C. (2004). Constructing and understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*, 79-151.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*, 1263-1277.

Charman, T., & Baron-Cohen, S. (1992). Understanding Drawings and Beliefs: a Further Test of the Metarepresentation Theory of Autism: a Research Note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 1105-1112.

Charman, T., & Campbell, A. (1997). Reliability of Theory of Mind Task Performance by Individuals with a Learning Disability: A Research Note. *Journal of Child Psychology and Psychiatry, 38*, 725-730.

Charman, T., Swettenham, J., Baron Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology, 33*, 781-789.

Colle, L., Baron-Cohen, S., & Hill, J. (2007). Do children with autism have a theory of mind? A non-verbal test of autism vs. specific language impairment. *Journal of Autism and Developmental Disorders, 37*, 716-723.

Costall, A., & Leudar, I. (2004). Where is the 'Theory' in Theory of Mind. *Theory and Psychology, 14*, 623-646.

Dissanayake, C., & Macintosh, K. (2003). Mind reading and social functioning in children with autistic disorder and Asperger's disorder. In V. Slaughter & B. Repacholi (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development.* (pp. 213-239). New York, NY: Psychology Press.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., et al. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology, 40*, 271-283.

Delinicolas, E. K., & Young, R. L. (2007). Joint attention, language, social relating, and stereotypical behaviours in children with autistic disorder. *Autism, 11*, 425-436.

de Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology, 24*, 7-37.

de Villiers, J. G., (2005). Can Language Acquisition Give Children a Point of View? In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 186-219). New York, NY US: Oxford University Press.

de Villiers, P. A., (2005). The Role of Language in Theory-of-Mind Development: What Deaf Children Tell Us. In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 266-297). New York, NY US: Oxford University Press.

Dissanayake, C., & Macintosh, K. (2003). Mind reading and social functioning in children with autistic disorder and Asperger's disorder. In V. Slaughter & B. Repacholi (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development.* (pp. 213-239). New York, NY: Psychology Press.

Dunn, J., Brophy, M., (2005). Communication, Relationships, and Individual Differences in Children's Understanding of Mind. In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 50-69). New York, NY US: Oxford University Press.

Fahie, C.-M., & Symons, D.-K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Applied Developmental Psychology, 24*, 51-73.

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2006). Specific Language Impairment, Theory of Mind, and Visual Perspective Taking: Evidence for Simulation Theory and the Developmental Role of Language. *Child Development, 77*, 1842-1853.

Flavell, J. H., & Miller, P. H. (1998). Social Cognition. In W. Damon, D. Kuhn & R.-S. Siegler (Eds.), *Handbook of Child Psychology* (5th ed., pp. 851-898). New York: John Wiley & Sons, Inc.

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive development (3rd ed.)*. Englewood Cliffs, NJ US: Prentice-Hall, Inc.

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 and Allied Disciplines, 46*, 409 - 419

Fodor, J. A. (1983). *The modulartiy of mind*. Cambridge, Mass: MIT Press.

Friedman, O., & Leslie, A. M. (2004). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science, 28*, 963-977.

Frith, U. (1989). *Autism: Explaining the enigma* Oxford: Basil Blackwell

Frith, C. D., & Frith, U. (1999). Interacting minds: a biological basis. *Science, 286*, 1692-1695.

Frith, U., & Happé, F. (1994). Autism: Beyond 'theory of mind.'. *Cognition, 50*, 115-132.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neuroscience, 14*, 433-438.

Glascoe, F. P., Martin, E. D., & Humphrey, S. (1990). A comparative review of developmental screening tests. *Paediatrics, 86*, 547-554.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*, 613-627.

Grant, C. M., Grayson, A., & Boucher, J. (2001). Using tests of false belief with children with autism: how valid and reliable are they? *Autism, 5*, 135-145.

Grant, C. M., Riggs, K. J., & Boucher, J. (2004). Counterfactual and Mental State Reasoning in Children with Autism. *Journal of Autism and Developmental Disorders, 34*, 177-188.

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6*, 514-523.

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

Harris, P. L., (2005). Conversation, Pretence, and Theory of Mind. In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 70-83). New York, NY US: Oxford University Press.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*, 233-259.

Hobson, P. (2004). *Cradle of thought: Exploring the Origins of Thinking*. Oxford: Oxford University Press.

Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good Test-Retest Reliability for Standard and Advanced False-Belief Tasks across a Wide Range of Abilities. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*, 483-490.

Hughes, C., (2005). Genetic and Environmental Influences on Individual Differences in Language and Theory of Mind: Common or Distinct? In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 319-339). New York, NY US: Oxford University Press.

Hughes, C., & Cutting, A. (1999). Nature, nurture and individual differences in early understanding of mind. *Psychological Science, 10*, 429-432.

Hughes, C., & Ensor, R. (2005). Executive Function and Theory of Mind in 2 Year Olds: A Family Affair. *Developmental Neuropsychology, 28*, 645-668.

Hughes, C., Jaffee, S., R., Happé, F., Taylor, A., Caspi, A., & Moffit, T., E. (2005). Origins of Individual Differences in Theory of Mind: From Nature to Nurture? *Child Development, 76*, 356-370.

Hutchins, T., L., Prelock, P., A., & Chace, W. (2008). Test-Retest Reliability of a Theory of Mind Task Battery for Children With Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities, 23*, 195-206.

Jarrold, C., Butler, D. W., Cottington, E. M., & Jimenez, F. (2000). Linking theory of mind and central coherence bias in autism and in the general population. *Developmental Psychology, 36*, 126-138.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32*, 70-78.

Joseph, R. M., & Tager-Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology, 16*, 137-155.

Kaland, N., Callesen, K., Moller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced theory of mind tasks. *Journal of Autism and Developmental Disorders, 38*, 1112-1123.

Kobayashi, C., Glover, G. H., & Temple, E. (2008). Switching language switches mind: Linguistic effects on developmental neural bases of 'theory of mind.'. *Social Cognitive and Affective Neuroscience, 3*, 62-70.

Leekam, S. R., Lopez, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology, 36*, 261-273.

Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition, 50*, 211-238.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*, 45-85.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225-251.

Leudar, I., & Costall, A. (2004). On the Persistence of the 'Problem of Other Minds' in Psychology Chompsky, Grice and Theory of Mind. *Theory and Psychology, 14*, 601-621.

Lewis, C., & Carpendale, J. (2002). Social cognition. In P.-K. Smith & C.-H. Hart (Eds.), *Blackwell Handbook of Childhood Social Development* (pp. 375-393). Cornwall: Blackwell Publishers Ltd.

Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development, 61*, 1514-1519.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., et al. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 261-270.

Lillard, A. S. (1993). Pretend play skills and the child's theory of mind. *Child Development, 64*, 348-371.

Lord, C., Rutter, M., DiLavore, P. S., & Risi, S. (1999). *Autism Diagnostic Observation Schedule - WPS*. Los Angeles: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659-685.

MacDonald, P. S. (2003). *History of the Concept of Mind: Speculations about Soul, Mind and Spirit from Homer to Hume*. Cornwall: Ashgate Publishing Limited.

Marzolf, D. P., & DeLoache, J. S. (1994). Transfer in Young Children's Understanding of Spatial Representations. *Child Development, 65*, 1-15.

Mayes, L., C., Klin, A., Tercyak, K., P. Jr, Cicchetti, D., V., & Cohen, D., J. (1996). Test-Retest Reliability for False-Belief Tasks. *Journal of Child Psychology and Psychiatry, 37*, 313-319.

Medrihan, L., Rohlmann, A., Fairless, R., Andrae, J., Döring, M., Missler, M., et al. (in press). Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses. *Journal of Physiology*.

Meins, E., & Fernyhough, C. (1999). Linguistic acquisitional style and mentalising development: The role of maternal mind-mindedness. *Cognitive Development, 14*, 363-380.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*, 622-646.

Morgan, B., Maybery, M., & Durkin, K. (2003). Weak central coherence, poor joint attention, and low verbal ability: Independent deficits in early autism. *Developmental Psychology, 39*, 646-656.

Moses, L. J., & Sabbagh, M. A. (2007). Interactions between domain general and domain specific processes in the development of children's theories of mind. In M. J. Roberts (Ed.), *Integrating the mind: Domain general vs domain specific processes in higher cognition.* (pp. 275-291). New York, NY US: Psychology Press.

Müller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development, 20*, 173-189.

Muris, P., Steerneman, P., & Merckelbach, H. (1998). Difficulties in the understanding of false belief: Specific to autism and other persuasive. *Psychological Reports, 82*, 51-57.

Muris, P., Steerneman, P., Meesters, C., Merckelbach, H., Horselenberg, R., van den Hogen, T., et al. (1999). The TOM test: A new instrument for assessing theory of mind in normal children and children with pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 29*, 67-80.

Nelson, K. (2005). Language Pathways into the Community of Minds. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind.* (pp. 26-49). New York, NY US: Oxford University Press.

Newborg, J. (2005). IL, USA Patent No. Riverside Publishing.

Newborg, J., Stock, J. R., Wnek, L., Guidubaldi, J., & Svinicki, J. (1984). IL, USA Patent No. The Riverside Publishing Company.

Noyes, J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: is workload another test mode effect? *British Journal of Educational Technology, 35*, 111-113.

Ontai, L. L., & Thompson, R. A. (2008). Attachment, parent-child discourse and theory-of-mind development. *Social Development, 17*, 47-60.

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*, 1081-1105.

Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders, 31*, 257-263.

Pears, K. C., & Moses, L. J. (2003). Demographics, Parenting, and Theory of Mind in Preschool Children. *Social Development, 12*, 1-20.

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: "Weak" central coherence and its relationship to theory of mind and executive control. *Developmental and Psychopathology, 18*, 77-98.

Pennington, B. F., & Ozonoff, S. (1996). Executive Functions and Developmental Psychopathology. *Journal of Child Psychology and Psychiatry, 37*, 51-87.

Peterson, C. (2002). Drawing Insight from Pictures: The Development of Concepts of False Drawing and False Belief in Children with Deafness, Normal Hearing, and Autism. *Child Development, 73*, 1442-1459.

Peterson, C. C. (2004). Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45*, 1096-1106.

Peterson, C. C., & Siegal, M. (1995). Deafness, Conversation and Theory of Mind. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36*, 459-474.

Peterson, C.-C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language, 15*, 123-145.

Peterson, C. C., & Wellman, H. M. (2009). From fancy to reason: Scaling deaf and hearing childrenâ€™s understanding of theory of mind and pretence. *British Journal of Developmental Psychology, 27*, 297-310.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in Theory-of-Mind Development for Children With Deafness or Autism. *Child Development, 76*, 502-517.

Peterson, D. M., & Bowler, D. M. (2000). Counterfactual reasoning and false belief understanding in children with autism, children with severe learning difficulties and children with typical development. *Autism, 4*, 391-405.

Premack, D., & Woodruff, G. (1978). Does the Chimpanzee have a 'theory of mind'? *Behavioral and Brain Sciences, 4*, 515-526.

Perner, J., Kloo, D., & Gornik, E. (2007). Episodic memory development: Theory of mind is part of re-experiencing experienced events. *Infant and Child Development, 16*, 471-490.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125-137.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*, 1228-1238.

Perner, J., Zauner, P., Sprung, M., (2005). What Does 'That' Have to Do with Point of View? Conflicting Desires and 'Want' in German. In Astington, J. W., & Baird, J. A. (Eds.) *Why language matters for theory of mind.* (pp. 220-244). New York, NY US: Oxford University Press.

Reddy, V., & Morris, P. (2004). Participants Don't Need Theories Knowing Minds in Engagement. *Theory & Psychology, 14*, 647-665.

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13*, 73-90.

Ruffman, T., Perner, J., & Parkin, L. (1999). How Parenting Style Affects False Belief Understanding. *Social Development, 8*, 395-411.

Russell, J., Hala, S., & Hill, E. (2003). The automated windows task: The performance of preschool children, children with autism, and children with moderate learning difficulties. *Cognitive Development, 18*, 111-137.

Sabbagh, M. A., & Seamans, E. L. (2008). Intergenerational transmission of theory-of-mind. *Developmental Science, 11*, 354-360.

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: An event-related potential study. *Psychological Science, 11*, 46-50.

Saxe, R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Research, 1079*, 57-65.

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience, 1*, 284-298.

Shanker, S. (2004). The Roots of Mindblindness. *Theory & Psychology, 14*, 685-703.

Shatz, M., & O'Reilly, A. (1990). Conversational or communicative skill? A reassessment of two year olds' behaviour in miscommunication episodes. *Journal of Child Language, 17*, 131-146

Slaughter, V., Peterson, C. C., & Mackintosh, E. (2007). Mind What Mother Says: Narrative Input and Theory of Mind in Typical Children and Those on the Autism Spectrum. *Child Development, 78*, 839-858.

Slaughter, V., & Repacholi, B. (2003). Individual Differences in Theory of Mind: What Are We Investigating? In B. Repacholi & V. Slaughter (Eds.), *Individual Differences in Theory of Mind Implications for Typical and Atypical Development* (pp. 1-12). New York: Psychology Press.

Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56*, 135-148.

Tager Flusberg, H. (2003). Exploring the Relationship between Theory of Mind and Social-Communicative Functioning in Children with Autism. In B. Repacholi & V. Slaughter (Eds.), *Individual Differences in Theory of Mind Implications for Typical and Atypical Development* (pp. 197-212). New York: Psychology Press.

Tager-Flusberg, H., & Joseph, R. M. (2005). How Language Facilitates the Acquisition of False-Belief Understanding in Children with Autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind.* (pp. 298-318). New York, NY US: Oxford University Press.

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Development, 79*, 284-302.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684.

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development, 75*, 523-541.

Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology, 44*, 618-623.

Williams, E. (2004). *Who* Really Needs a 'Theory' of Mind? *Theory and Psychology, 14*, 705-724.

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

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and 'false' photographs. *Cognition, 35*, 41-68.