DEVELOPING THEORY OF MIND AND EXECUTIVE FUNCTIONS FROM THREE- TO FIVE-YEARS-OLD: CROSS-SECTIONAL GROUP AND LONGITUDINAL SINGLE-CASE APPROACHES

by

SARA THERESA BAKER

A Dissertation submitted to the

Graduate School- New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in the Department of Psychology

Written under the direction of

Alan M. Leslie

And approved by

________________________________________

________________________________________

________________________________________

________________________________________

New Brunswick, New Jersey

October, 2007

ABSTRACT OF THE DISSERTATION

Developing theory of mind and executive functions from three- to five-years-old:

Cross-sectional group and longitudinal single case approaches

By SARA THERESA BAKER

Dissertation director:

Alan M. Leslie

In this dissertation I consider traditional approaches to developmental questions, and suggest new methods for analyzing variability in individual children. In chapter I, I review the literature on children's reasoning about mental states, such as beliefs and desires, from a very early age. This ability is often called 'theory of mind'. In chapter II I explore the suggestion that change in a preschooler's theory of mind is motivated by change in their executive functions, which reflect the child's ability to manipulate abstract representations. To this end, a new measure of cognitive development in executive functions is tested empirically. A correlation between theory of mind performance and executive function performance is demonstrated. In chapter III, I go beyond correlations in groups, to address the question of individual development. Current theories of how theory of mind develops do not make any specific predictions about what developmental change looks like in individuals. Is change abrupt or is it gradual? Is there a universal path to mature performance, or are there inter- and intra-individual differences? A new method for statistically analyzing variability in individual children's performance is proposed. Findings are discussed in terms of a general theoretical framework for theory of mind development in chapter IV.

Acknowledgements

First and foremost, I would like to thank the children, parents, and teachers who graciously gave their time for this research. They made this work possible, and worthwhile. I am indebted to Alan Leslie for his careful mentoring throughout my graduate career. Ori Friedman, Randy Gallistel, Rochel Gelman, Stathis Papachristos, and Henry Wellman gave priceless comments on both methodological and theoretical aspects of the work. It should not go unmentioned that the Center for Cognitive Science at Rutgers University was a friendly and inspiring place to study. Interdisciplinary discussions made our new approaches possible. My dissertation work was facilitated by talented research assistants in the Cognitive Development Lab at Rutgers, to whom I am grateful. And finally, my friends and family deserve a sizeable share of gratitude for their support and encouragement. This work was supported in part by National Science Foundation grants to AML (BCS-0079917 and BCS-0725169).

Table of contents

List of tables

# I. Introduction

I review the classic literature on children's theory of mind and make a case for a developmental relation between theory of mind and executive functions, which is more than mere correlation. To this end, I describe the few empirical findings that exist showing how this relation might play out in individual children, who are, after all, the locus of development. New methods for investigating inter- and intra-individual variability are proposed.

## I.I. Children's theory of mind in the classic literature

Classically theory of mind development is assessed by telling the child a short story with the aid of toys or pictures (see figure 1). At the end of the story, children are required to predict what will happen next, based upon the character's beliefs and feelings (rather than their own). For example, children see a doll named Sally put her ball in a basket while her friend Anne watches. Then, after Sally leaves the room, Anne transfers the ball to another location, say, a box. Given that Sally did not witness this unexpected transfer, children are asked to predict *where Sally thinks the ball is*. If they answer that Sally thinks the ball is still in the box, then they are considered to have passed the test of theory of mind. Predicting that Sally has a false belief about the location of the ball demonstrates the ability to attribute beliefs to Sally on the basis of her informational access, even when Sally has not seen everything the child has seen.

The non-overlapping access to information between Sally and the child, created by the absence of Sally during a crucial unexpected transfer event that the child witnesses, is necessary in order to establish that answers are given based upon Sally's mental states, and don't simply happen to coincide with the child's own mental states. Children are also asked a set of control questions about the storyline, to ensure that their answers stem from accurate memory of the events, and any failure is not due to forgetfulness.

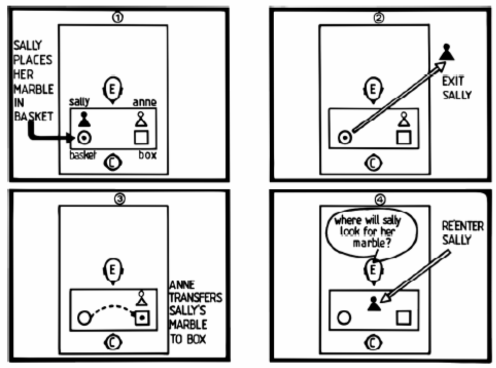



Figure 1. The classic unexpected transfer test of theory of mind (Baron-Cohen, Leslie and Frith, 1985).

**I.I.1. Classic developmental timeline of false belief attribution tasks**

In the original Wimmer and Perner (1983) study, only 50% of the children between four-and five-years-old could properly attribute false beliefs to the dolls. All of the six- and seven-year-olds passed the belief attribution test, with only one exception. Since the inception of the false belief paradigm, there have been countless studies of children's developing theory of mind, both in normal and disabled populations (Baron-Cohen, Leslie & Frith, 1985; Hogrefe, Wimmer & Perner, 1986; Perner, Leekam & Wimmer, 1987; Leslie & Thaiss, 1992; Clements & Perner, 1994; Roth & Leslie, 1998; Hughes, 1998; Moore, Pure & Furrow, 1990; Leslie & Polizzi, 1998; Friedman & Leslie, 2004; for a review, see Wellman, Cross & Watson, 2001).Through simplifications of the original scenario, in terms of the verbal load and the events the child witnesses, the passing age has been pushed back from six-years-old to four-years-old. Still, three-year-olds typically fail to predict where Sally thinks the ball is.

Interestingly, group performance at three-years is not random. At 30 months of age, 80% of children predict that Sally will look for the ball in its current location, as if the doll had seen what they have seen, even though she was absent (Wellman, Cross and Watson, 2001). It is not the case that three-year-olds choose an answer at random. On the contrary, young three-year-olds overwhelming tendency is to predict that the character's belief will be the same as their own belief. Young preschoolers tend to attribute beliefs to Sally which she would have no reason to hold, given what we know about her access to information, as if it is too difficult for the young child to overcome their own perspective on the problem. This aspect of their failure, as a group, to demonstrate theory of mind will be important to consider when we need to offer an account of why children fail and why they might pass these tasks.

Group data show that by the age of five, most preschoolers correctly predict that Sally thinks the ball is still where she put it before leaving, demonstrating that they can take the character's perspective despite it being different from their own (Baron-Cohen, Leslie & Frith, 1985; Perner, Leekam & Wimmer, 1987; Leslie & Thaiss, 1992; Clements & Perner, 1994; Roth & Leslie, 1998; Hughes, 1998; Moore, Pure & Furrow, 1990; Leslie & Polizzi, 1998; Friedman & Leslie, 2004; for a review, see Wellman, Cross & Watson, 2001).

In a control task where Sally's belief about the ball is true because she returns to witness the ball being transferred to the box, group data show three-year-olds typically understand that Sally knows where the ball is. So difficulty for young preschoolers depends upon the non-overlapping points of view held by the child and the character in the story.

### I.I.2. Converging evidence for false belief attribution in preschoolers

Another classic test of theory of mind requires children to report someone's belief about the unexpected contents, rather than the unexpected location, of an object (Hogrefe, Wimmer & Perner, 1986; Perner, Wimmer & Leekam, 1987; Perner, Frith, Leslie & Leekam, 1989; Gopnik & Astington, 1988). In this theory of mind task, children are first shown a familiar container, say, a candy tube. After the child says what they think is inside (Candy), the experimenter opens the tube to reveal an unexpected contents, say, a pencil. Children are then asked to report their own prior false belief, that is, what they first thought was in the container (Candy). Then, children are asked to report what someone else would think was in the container (Candy). The classic finding from this test parallels the one from the unexpected transfer test with Sally and Anne. Young preschoolers fail to report false beliefs. Instead, they tend to report that they first thought the tube contained whatever they now know it contains (demonstrating the well-known *hindsight bias*). Also, young preschoolers predict that someone else will think the tube contains the unexpected, rather than the expected, contents, as if the other person has access to the knowledge they now hold. Both cases require the child to report a belief that is different from the one they currently hold to be true. In a similar test of false belief attribution, Flavell and colleagues (1986 for a review) systematically demonstrated what they called a "phenomenism error": three-year-olds' inability to say that an object that looked like a rock was really and truly a sponge, after they had gained knowledge about the actual identity of the object. The youngest preschoolers persisted in reporting the appearance of the object ("it's a rock"), rather than the actual identity ("it's a sponge").

Performance on these classic tests of theory of mind measuring children's ability to report someone's false belief are moderately correlated, even when age and verbal ability are accounted for, suggesting a fair degree of inter-task consistency (Hughes, 1998; Wellman, Cross & Watson, 2001).

One great motivator for early research on theory of mind was that autistic children, even through their teens, have social deficits. In particular, they seem to have trouble reasoning about mental states. Autistic children fail this belief attribution task, while children with other developmental delays, such as Down's syndrome, do not (Baron-Cohen, Leslie & Frith, 1985). I later return to the topic of autism when I consider the domain-specificity of theory of mind.

As far as the basic findings I just reviewed are concerned, much research has accumulated demonstrating the universality of preschoolers' developing theory of mind (see for example Sabbagh, Xu, Carlson, Moses, & Lee, 2006; Pears & Moses, 2003; Hughes, & Cutting, 1999). Across many cultures and educational, biographical and socio-economic contexts, children have demonstrated similar developmental timelines, leading researchers to conclude this sort of social cognition depends upon innate capacities for recognizing and reasoning about people's thoughts and feelings.

**I.II. Operations of mental state reasoning: Models and Predictions**

The field currently hosts at least two competing views of how belief attribution works, theory-theory and modular-theory. These two accounts of mental state reasoning differ in the constitutive operations by which concepts are acquired and updated in everyday life. Though there are others, I only discuss two positions here.(1)

**I.II. 1 Modelling theory of mind: Two views**

According to theory-theory, individuals use hypothesis-testing to develop a general theoretical understanding of mental states (Schulz & Gopnik, 2004). There are three features of a theory, which are present from the beginning in rudimentary form in a child's developing ability to understand people as mental agents and which liken their understanding to a "theory" of how the mind works (Wellman, 1988). First, the theory is comprised of a coherent body of interconnected concepts, such as beliefs, desires, intentions and motivations. Second, the theory is about a circumscribed body of phenomena, constituting an ontology proper to the theory. That is, the extension of a theory is limited to one domain of knowledge. For theory of mind, it is the domain of naïve psychology. Thus, there would be a separate theory for the domain of naïve physics. And third, a theory provides a general causal framework following which events in the world can be explained. Theory-theory holds that the same hypothesis-testing method is used for reasoning about causal and factual events in the world as for one's own mental states and a third person's mental states. In other words, the means by which children acquire a theory-like understanding of mental states is not a domain-specific mechanism, but a general-purpose learning (theory-building) mechanism.

A separate account of theory of mind development is the modularist view. The modularist view holds that reasoning about mental states uses domain-specific mechanisms. This contrasts with classic theory-theory whereby a general inferential mechanism (e.g., hypothesis-testing) is used across domains (such as naïve biology and naïve physics). According to modularism, when we reason about mental states a specialized theory of mind mechanism (ToMM) first identifies possible propositional contents of the mental state (such as the true belief location or the false belief location). ToMM is assisted by a selection processor (SP) which weighs the candidate contents for the mental state, then selects among the candidates a target for attribution (Leslie, German & Polizzi, 2005).

**I.II.2 Developmental predictions**

Notwithstanding the differences between these models' explanations of how concepts of mental states are acquired, we can't help but notice they have something in common. Neither model makes specific predictions about the profiles of development we could observe in individuals. Both of these models are consistent with developmental curves that change smoothly or abruptly, and both of these models are consistent with monotonic as well as non-monotonic change. What both of these models are neutral about is crucial to developmental science: a) a description of the developmental path, including rate of change, b) an explanation of development in terms of the motivating factors which influence change. Developmentalists have wrestled with these general questions for a long while, not just in theory of mind research. Simply put, whether theory or module is the means for acquiring and using concepts of mental states, and whether the child's concepts undergo radical qualitative change, or change is quantitative, we must explain children's ability to have representations of mental states, and manipulate them in a context-appropriate fashion (where the context is defined not only by physical environment, but also by the epistemic environment).

### I.III. The case for an executive component in theory of mind reasoning

The ability to manipulate representations in a context-appropriate fashion is often attributed to executive functioning, a term loosely referring to a subject's ability to plan or update a response, given a novel situation or strong urge to give an incorrect answer.

Early in the literature on children's ability to reason about mental states it was suggested that their difficulty stems from trouble manipulating discordant representations (Russell, Mauthner, Sharpe, & Tidswell, 1991). In their paper, these authors showed that children who were unable to perform well in a false belief task were also unable to overcome their natural tendency to point to a chocolate, in a game where only pointing to the empty box could win the chocolate. Preschool children in the Russell et al. study were first trained to point to one of two opaque boxes. The child and the experimenter-competitor were both blind as to what the boxes contained. If the child chose a box which contained a chocolate, the experimenter won the treat. If it was empty then the child won the treat. This initial “blind” phase served to show the children what the results would be of pointing to an empty box (= they won a treat) versus pointing to a box with chocolate in it (= the competitor won the treat). Next, the same task was administered but with boxes which only the child could see through. These “windows” allowed the child to calculate the best answer on each trial, and point to the box they knew was empty. Now the children would need to use their best skills of deception to win chocolates, namely, pointing to a box they could see was the empty one. They were given 25 trials. Three-year-olds found this difficult (only one out of 17 passed on the first test trial and only six out of 1 passed on any of the trials), while four-year-olds had less trouble being deceptive (10 out of 16 passed on the first trial and they all passed at least 18 trials). The authors calculated a correlation between the preschooler’s performance on the false belief task and their performance on this new “windows” task, and showed that there was a significant relation. This led them to conclude preschool children might have trouble in reporting someone else’s false belief because they fall victim to the salience of their own knowledge and cannot easily manipulate their own current epistemic states (as demonstrated by the young preschoolers’ tendency to point to the box where they know the chocolate is, despite repeatedly experiencing the loss of the treat). Calling this the salience hypothesis, Russell et al. were careful to explain that they were not suggesting children were simply impulsive. Their suggestion went beyond a behavioral (or response) difficulty, and targeted children’s ability to begin to conceive of alternative representations when they had access to one very salient one (usually because it stems from a current state of reality). It is crucial to remember that it is not a problem for three-year-olds in these paradigms to report *just any past states* which conflict with current states, it is a problem for them to report past *mental states* which conflict with their own current mental states. This is why Russell and colleagues took pains to explain their position. According to their account, having a theory of mind means that children are able to inhibit salient knowledge and overwrite knowledge of *epistemic* states, or in other words, having a theory of mind means that they exhibit executive control.

Indeed, there is reason to believe that autism is in sum an executive dysfunction (Russell, Mauthner, Sharpe & Tidswell, 1991; Hughes & Russell, 1993), and some researchers still support the thesis that reasoning about mental states is simply equitable to reasoning about complex propositions (Frye, Zelazo, & Muller, 1995). So let us review the case for the developmental relation between theory of mind and executive functions.

**I.III.1. Overview of executive functions in cognitive development**

Executive functions reflect the exercise of context-appropriate behaviors, regardless of one's reflexive response. Historically, executive functions have been linked with the prefrontal cortex. Luria (1973) famously developed a battery of tests for adult patients with frontal lobe damage, and showed that these patients generally had trouble overcoming the pre-potent response to a stimulus. For example, they could not apply even a simple rule like "point your finger when I make a fist, and make a fist when I point my finger". Instead, patients with frontal lobe damage tend to respond following their first impulse, in this case matching the movement made by the examiner.

Children have recently become the focus of investigations of developing executive functions, as researchers develop more ways to test executive competence. There exist a host of tests for preschoolers' executive functions, which I will not discuss (e.g, noisy book task, Hughes, 1998; tower of London, Shallice, 1988). Here I limit our discussion to those tests which reliably show a correlation in the preschool years with developing theory of mind. The first one requires children to give a verbal response when they see a picture on a card. The rule states that they should say “day” when they see a picture of the moon and “night” when they see a picture of the sun. Because they must overcome the pre-potent tendency to say what they see, young preschoolers without mature executive functions to apply the contextual rule will fail to produce the correct answer. Reminding children of the rule does not seem to help their performance. In fact, it seems that three-year-olds have trouble even recognizing that they are wrong, whereas children a bit older may give the wrong answer but correct themselves, or show some outward sign of underlying competence, despite the continuing difficulty in producing the correct responses. By the age of five, children generally do not find this task difficult, saying the correct and opposing word each time they see a card, almost without hesitation.

Another test of preschoolers’ executive functions requires them to sort cards into boxes according to one rule, then to sort according to an orthogonal rule. Imagine a deck of cards showing two types of objects (e.g., a car and a dog) in two different colors (e.g., red and yellow). The deck is comprised of four types of cards. In the first phase, the child is asked to sort the cards along one dimension (say, color). This is easy even for three-year-olds. In the second phase, the child must switch rules, and sort the cards along the second dimension (shape). It is difficult for young preschoolers to apply the second rule over the same set of cards used in the first task. They commit an error of perseveration, continuing to use the first rule they were asked to use. They do this despite being able to tell you where the cards should go, so again, it is not for lack of memory for the current rule that young preschoolers fail. When it comes to the second phase after the rule switch they simply seem to have trouble overcoming the strong response instilled in the first phase. By the age of five, children easily sort the cards after the rule has changed, demonstrating stronger executive functions.

**I.III.2. Correlations between executive functions and theory of mind in the preschool years**

The correlations between children's performance on these executive measures and performance in classic theory of mind tests suggests there may be a developmental relation between these two types of competence (Carlson & Moses, 2001; Carlson, Moses & Breton, 2002; Perner, Lang & Kloo, 2002; Russell, Mauthner, Sharpe, & Tidswell, 1991; Perner & Lang, 1999; Hughes, 1998; Flynn, O'Malley & Wood, 2004). Consistently, researchers have found moderate positive correlations between performances on these tasks. Perner, Lang and Kloo (2002) reported a strong correlation between prediction in a False Belief-location task ('Sally/ Anne') and inhibition as measured by the Card Sorting task ($r = .65$). Carlson and Moses (2001) tested 107 three- and four-year-olds on a battery of ten Inhibitory Control tasks (e.g., Day/ Night, Card Sort, Gift Delay, Whisper) and a battery of four theory of mind tasks (False Belief-location, False Belief- contents, Deception and Appearance- Reality). They also found that overall, the theory of mind battery correlated with the Inhibition battery ($r = .66$). Inhibitory processing itself can be further divided into (at least) two sub-components: inhibition of competing responses (termed 'Conflict' inhibition by Carlson and Moses, 2001) versus inhibition of one response ('Delay' inhibition).

Grouping the inhibition measures under two headings, 'conflict' tasks involving competing answers of which only one should be selected, and 'delay' tasks involving a waiting period before the response, the authors found a higher correlation with ToM tasks for the 'Conflict' battery ($r$ = .64) than for the 'Delay' battery ($r$ = .49), although both correlations were statistically significant. This was not fully replicated in a later study. Carlson, Moses and Breton (2002) further examined the distinction between 'Conflict' inhibitory demands and 'Delay' inhibitory demands. The results here diverged from the previous study. This time, children's performance on theory of mind tasks, Appearance- Reality and False Belief, correlated significantly with the 'Conflict' inhibition measures ($r$ = .30 and $r$ = .52, respectively), but not the 'Delay' inhibition measures ($r$ = .19 and $r$ = .17, respectively). The question raised by diverging results of this nature is that of other factors that may be affecting performance. One way to investigate this is partial correlations.

As with any correlational results, one might wonder whether they are due to other common factors. In Carlson and Moses (2001), performance on the theory of mind battery and the Inhibitory Control battery remained significantly correlated even when age, gender and verbal ability were held constant ($r$ = .41). Also, when age, gender and verbal ability were partialled out, and when the Inhibitory Control tasks were split between 'Delay' and 'Conflict' tasks, there remained a significant correlation with ToM ($r$ = .23 and $r$ = .41, respectively). Compare once again with Carlson, Moses and Breton (2002). This time, using partial correlations, the authors controlled for age and IQ. The only remaining statistically significant correlation was between False Belief and 'Conflict' measures ($r$ = .30). Appearance- Reality no longer correlated with either 'Conflict' measures or 'Delay' measures, and False Belief no longer correlated with 'Delay'. To make matters even less clear, Perner, Lang and Kloo (2002) found that holding age, verbal ability and control question performance constant yielded non-significant, at times negative, correlations between False Belief- location scores and other inhibition tasks (Card Sorting, Go/ No-Go) and ($r$ = .14 and $r$ = -.08, respectively). So there exists a set of studies that do not speak in one voice about the correlation between theory of mind and executive functions, when other factors, presumed to influence performance, are held constant. What does this say?

### I.III.3. Interpreting the correlational evidence

Interpreting these variable results, which are obtained by comparing groups of different ages on tasks they complete just one time, is made difficult by diverging procedures. First, the factors researchers chose to partial out of their analyses were different for each of the studies (Carlson & Moses, 2001: age, gender and verbal ability; Carlson, Moses & Breton, 2002: age and IQ; Perner, Lang & Kloo, 2002: age, verbal ability and control question scores). Second, the questions used, and therefore the tasks themselves, although very similar from an adult's perspective, may lead to various levels of difficulty for children who are chosen precisely because they are in a critical time-window of development for the ability under investigation. This means that the slightest modification from one procedure to the next may induce noise. Notably, in the studies I discuss here, there are serious differences in the way they assessed theory of mind understanding for the classic unexpected transfer false belief task ('Sally/ Anne'). Perner, Lang and Kloo (2002) asked the children *where Sally will first go* to get her ball, Carlson and Moses (2001) asked the children *where Sally thinks her ball is*, and Carlson, Moses and Breton (2002) asked the children *where Sally thinks her ball is* in one version, and *where Sally will first look* for her ball, in another version. This particular manipulation, the addition of the word 'first' to the critical question, has been shown to affect children's performance on the Sally/ Anne task, as described later, and therefore renders a comparison of these correlational studies more difficult, especially given that these authors claim ‘look first’ reduces inhibitory demands (Surian & Leslie, 1999; see also Leslie, German, & Polizzi, 2005). What can be said with a reasonable degree of confidence is that there is some sort of relationship between developing executive functions and developing theory of mind. But what kind of relationship are we dealing with?

**I.IV. What is the nature of the developmental relation between executive functions and theory of mind?**

The relation in question may present five forms: 1) theory of mind and executive functions may not be correlated after all, that is, they may not share any causal developmental relation; 2) theory of mind and executive functions may be correlated because they are in fact the same capacity, described in two different ways by theorists; 3) theory of mind and executive functions may be correlated because they rely on a common underlying factor (for example, they could recruit partially overlapping brain regions); 4) theory of mind and executive functions may be correlated because developing executive functions rely causally on developing theory of mind; and finally, 5) theory of mind and executive functions may be correlated because developing theory of mind relies causally on developing executive functions. I now discuss each of these options in turn and find that I can already rule out several of them based upon empirical evidence.

**I.IV.1. Is there really any relation between executive functions and theory of mind development?**

First, I attempt to rule out the possibility that theory of mind and executive functions are not meaningfully related to each other in development. For this, I take evidence from two task manipulations which demonstrate that changing the executive requirements of theory of mind tests affects children's performance. In one manipulation, children were shown a story very similar to the classic unexpected transfer story. However, children never saw the ball in any hiding location; they were only told about it (Zaitchik, 1991). As in the classic test, children had to report someone else's belief when it was different from their own, but they did not have to overcome their own knowledge of the ball's location to do so. This manipulation made the task slightly easier for young preschoolers. That is, the young three-year-olds in the study found it easier to report someone else's belief when their own belief was rendered more tenuous, and the representation of the ball's actual location was weaker in their own minds. The effect of this manipulation can be interpreted as reducing the salience of the child's own belief, thereby making it easier to report someone else's differing belief.

Along similar lines, another manipulation of the classic unexpected transfer task yielded better performance in three-year-olds as compared to the standard way of testing their theory of mind understanding. This time, researchers altered the protocol minimally, simply adding the word "first" to the prediction question. Now children were asked: "Where will Sally look *first* for the ball?". Siegal and Beattie (1991) and Surian and Leslie (1999) found that adding the word 'first' to the critical question helped three-year-olds attribute a false belief to Sally.

The intention was to test whether it would help children to place a greater emphasis on Sally's belief, rather than their own. Again, manipulating the salience of the character's belief relative to the child's own belief helped them. In both cases, when the character's false belief was stressed, either by raising its salience or by reducing the salience of the child's true belief, children's performance benefited. Taken together, these two empirical lines of evidence seem to suggest we can rule out the possibility that theory of mind and executive processing are not related in a meaningful way. But in what way are they related? Could it be that these two theoretical constructs are in fact one and the same?

**I.IV.2. Could theory of mind and executive functions be the same thing?**

Autistic children do appear to have deficits in both theory of mind and executive functions, and normal children's abilities in these two domains are correlated. A parsimonious appraisal might suggest that the theoretical constructs of theory of mind and executive functions operate on the same underlying neuro-cognitive mechanism. This is closely related to Russell and colleagues suggestion that in essence, being capable of recognizing and reasoning about mental states is equivalent to utilizing executive functions. However, there are compelling results which indicate that theory of mind and executive functions are not reducible to the same abilities in children.

**I.IV.2.1. A double dissociation on representational change tasks**

The prospect of collapsing theory of mind and executive functions was distanced when researchers compared autistic children, normally developing preschoolers and children with Down syndrome on a series of tasks (Leslie & Thaiss, 1992). The battery was designed to be equal in executive requirements so that any differences in performance could be attributed to the tasks' varying requirements for reasoning about mental states. In the classic unexpected transfer test children were asked to predict Sally's false belief. Her belief is an outdated representation of reality. Additional tasks were used, wherein the outdated representation of reality was a photograph, a map, or a drawing, rather than a belief. These physical (as opposed to mental) representations could be outdated as well, by changing something in the scene once the representation is created and placed out of sight. The authors reasoned that if theory of mind is reducible to the ability to manipulate representations, then maintaining and reasoning about outdated representations in any format (be it a belief, a photo or a drawing) should elicit the same patterns of performance. In fact, they uncovered a double-dissociation between the mental state tasks and the analogous tasks, when comparing autistic and normal children's performance (see figure 2).

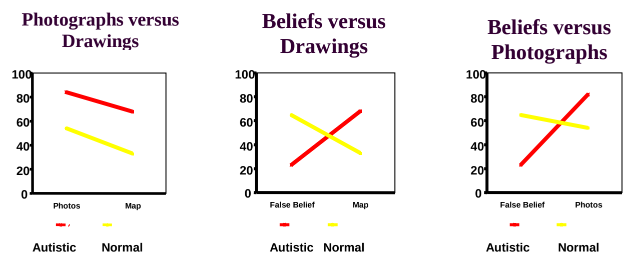



Figure 2. Comparison representational change tasks between populations (Leslie & Thaiss, 1992)

Autistic children performed well when they were asked to describe an outdated representation in the photographs and maps versions, compared to normal three-year-olds. However, when it came to describing outdated mental representations like Sally's belief in the story, autistic children performed well below normal.

This double dissociation in autistic and normal performance is informative insofar as it sheds light on the developmental relation between theory of mind and executive functions. Three-year-olds and autistic children are two populations who typically fail classic tests of false belief understanding, but for different reasons. Three-year-olds seem to fail because they lack executive resources for overcoming the dominant candidate for belief attribution (i.e., their own true belief). Autistic individuals seem to fail a test of false belief understanding because they lack the understanding of how the mind works which would lead them to identify a belief candidate amongst many possibilities. That is, autistic individuals apparently fail to appreciate mental states, reasoning instead about the world as if there are only representations like photos or maps.

Because these tasks were analogous in their executive requirements, the difference in performance between normal and autistic children is attributed to a domain-specific component in theory of mind reasoning. Understanding mental states is more than simply being adept at manipulating competing representations. It requires in addition that children be able to identify possible belief candidates that are distinct from their own currently held beliefs. It is this ability that seems to lag behind in autism, and not a domain-general ability to manipulate representations.

**I.IV.2.2. Autistic and normal performance is not influenced by the same factors**

The striking differences in between normal theory of mind and autists' theory of mind do not stop there. Results showing facilitation of false belief attribution for normally developing three-year-olds, when their attention is directed towards the false belief location and away from the true belief location, were not obtained with autistic children. In their 'look first' manipulation, Surian and Leslie (1999, experiment 2) showed that increasing the salience of the false belief location did not help autistic children pass. They continued to fail to attribute a false belief to Sally. Not one of the autistic children who were tested passed the 'look first' question while failing the standard version ('first' not pronounced in the question), whereas 30% of the three-year-olds who were tested failed the standard question and passed the 'look first' question. Manipulating the executive requirements in the Sally/ Anne task may help normal children report someone else's false belief, but it cannot give autistic children access to understanding of mental states. Autistic children must lack a domain-specific component to mental state reasoning. This comparison of autistic and normal performance shows once again that not only is there a developmental relation between the strength of executive functions and a child's theory of mind, but also that these two sets of competence cannot be collapsed into a single theoretical construct.

**I.V. Various approaches to the developmental problem of interacting processes**

The picture of cognitive processes underlying successful theory of mind reasoning is becoming clearer. I have ruled out the possibility that theory of mind and executive functions are unrelated, as demonstrated by the effects on normal children's performance when the executive requirements are modified within mental state reasoning tasks. I have examined the possibility that theory of mind and executive functions are reliably correlated in preschool because they reflect the same underlying ability. This possibility has been distanced by the striking double dissociation between autistic children's performance and normal children's performance on analogous representational change tasks. This conclusion was strengthened when I observed that autistic children are not helped by the same task features that normal children might be. In our search to determine what developmental relation holds between theory of mind and executive functions, I leave open the possibility that correlations in cognitive functions appear because they rely on shared brain regions. This could be the case regardless of what we decide about a *cognitive* functional dependence between the two types of competence.

**I.V.1. What does the norm tell us?**

Thus far I have reviewed the classic evidence on theory of mind in normally developing preschoolers and in autistic children. I have also presented an argument to support a developmental relation between theory of mind and executive functions, which goes beyond a simple coincidence. I am still hoping to elucidate the nature of this relationship, and offer a causal explanation as to why certain manipulations of the experimental protocols improve performance for one population and not another. Although the cross-sectional evidence supports the possibility of causal dependence between theory of mind and executive functions, I have not yet left the realm of group data. Cross-sectional group averages can be difficult to interpret when it comes to describing dynamic processes, because they capture only a static picture and may reflect artifacts of averaging.

I believe that it is essential to go beyond the plotting of curves of normal behavior, and ask, for any individual within the curve, what it is about the mechanisms of reasoning they employ which places them at that point of the curve. This sort of approach to development will encourage research that brings out causal explanations of mechanisms of change, which are to be preferred over a succession of descriptions of states with no way of accounting for a switch from state *a* to state *b*. Neither theory-theory nor modular theory makes any specific predictions as to the developmental profile of theory of mind within an individual child (is change abrupt or is it gradual?). I propose to study just that.

**I.V.2. A meta-analysis of group averages in theory of mind**

If one browses the group data generated by the host of studies world-wide, as Wellman and colleagues did in their 2001 meta-analysis, one finds an *apparently smooth increase in performance on these tasks* (Wellman, Cross & Watson, 2001; see figure 3). Given that children's answers could be either right or wrong, a monotonically increasing function is classically interpreted as representing a series of abrupt changes within individuals, smoothed out through group averages.

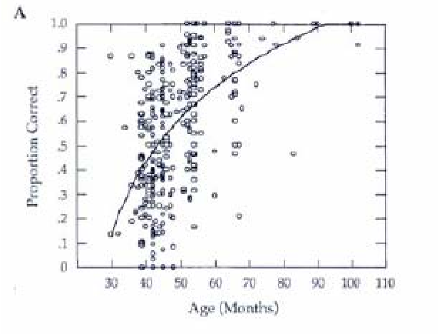



Figure 3. Scatterplot of group averages in several theory of mind experiments (Wellman, Cross & Watson, 2001)

Assuming that individual behaviors through a period of change can be represented by group averaged curves is not reserved for child development. As a matter of fact, researchers in animal learning study acquisition behaviors in a variety of ways, often involving frequent sampling of their behavior during a variety of learning tasks. Traditionally, models of acquisition in learning theory have represented animal learning as a steady monotonic process. Under this view, if the group-average curve is a good representation of what the individual learning curve looks like, mice should gradually adjust their poking to match the probability that food appears in a given hole. In a situation where only one hole rewards the mouse's poke with food, the mouse should eventually poke only in that hole, and they should never return to poke in the other hole yielding no food. This would fit the classic assumption of monotonic improvement in a task.

**I.V.3 Individuals' paths of change cannot be captured in group averaged curves**

However intuitive this model of developmental change may be, this seems less and less likely in light of the data. Investigations of individual mice's learning curves reveal a variety of patterns, most of which are in fact non-monotonic (Papachristos & Gallistel, 2006). Reversal of appropriate behavior may occur during learning, because the animal is not yet sure of the correct answer. More surprising, though, is what the authors observe *after* the mice have reached the asymptotic performance. Almost half of them (9/ 20 individual mice) still show statistically significant reversals in their behavior at some point after they have learned their task (as calculated by their responding near asymptote; see figure 4).

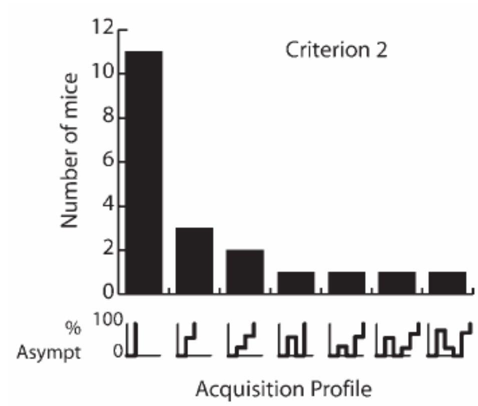



Figure 4. Individual profiles of acquisition (Papachristos & Gallistel, 2006)

The human equivalent of the head-poke in our theory of mind task is the prediction the child makes to our question about Sally's belief. If the traditional model of learning applies to child development, then children should show a steady improvement in performance, at some point switching their answers from the incorrect prediction to the correct one, and never reverting back to the incorrect response. Indeed, the classic picture of theory of mind development describes a progression occurring around the age of four from an early immature perception of people's thought and feelings, towards a more mature vision of how these mental states interact to predict behavior. The classic picture is based upon the universal findings described above on the unexpected transfer and the unexpected contents tasks. These young preschoolers do not seem to be guessing, because they are almost always wrong about their predictions. Group data show that children near the fourth birthday no longer consistently fail, but they do not yet consistently pass either. After four, children are capable of predicting a false belief. But again, is this what happens within each individual child who develops their theory of mind? Putting this developmental program in terms of children's ability to use mental states in their understanding and prediction of actions, we should investigate what motivates developmental change, that is, where difficulty stems from, and what changes occur to eventually allow this difficulty to dissipate. Why is it very difficult for a young three-year-old to predict when someone has a false belief, but easier for a four-year-old?

Classic studies use a cross-sectional approach to developmental change, as it can reveal differences between average performance at different ages. However, this cross-sectional approach assumes a monotonic change in each individual's understanding of the problem, such that a group average at time *a* and a group average at time *b* would be worth comparing, as in "pre" and "post" change. But is this really what happens when children develop an understanding of mental states? Is it possible that the nature of developmental change is more complex than mapping out a "before" and "after" and putting these static pictures side by side? Beyond the intra-individual variability implied by the possibility of non-monotonic change, we should also consider the possibility of various patterns of development across individuals. Both of these types of variability are ignored in group averages, and I suggest they deserve more attention from developmental researchers.

### I.V.4. Towards a method for capturing the dynamics of development

Our models of learning and development based on group averages, giving static "before" and "after" pictures of states, can not capture the path of the developing processes. I take this to mean that the classic approach to describing the development of children's understanding of mental states with group averages is insufficient. If we hope to explain where children's difficulty comes from, and how they ultimately demonstrate their competence, we must look at development within the individual child.

As I have already said, when looking to dress a model of the development of theory of mind in children, I will need to consider which sorts of manipulations in the task have helped or hindered children's performance. This may give us clues as to what processes are more or less mature, and put us on a path towards accounting for the changes throughout development. I am aware of only a few studies which look at the *same children over time* in order to uncover developmental trends.

## I.VI. Capturing variability: Following the same children through the course of development

### I.VI.1. Longitudinal sampling of theory of mind development

In the search for a way to further our understanding of the developmental relation between theory of mind and executive functions, I now turn away from group average data, towards methodology using long-term data for comparison of behaviors within and between children. More studies are being undertaken with the aim of studying development *throughout* a period of change- namely, longitudinal studies and microgenetic studies.

Very few truly developmental studies exist which address the question of causal mechanism underlying developmental change in theory of mind. The first types of studies are longitudinal in nature, with the aim of examining children's competence over a period of time. Claire Hughes (1998) asked whether children's performance on one task at time *a* predicts their performance on other tasks at time *b*. She first examined three- and four-year-olds performance on classic theory of mind tasks (Sally/ Anne prediction and explanation, and Candy prediction for self and explanation for other) and executive function measures (detour-reaching box inhibition task, Luria's hand game inhibition without action plan elaboration, total-change set-shifting, noisy-book working memory task). Then, after 13 months, she re-tested the same children on those same tasks, and a few additional tasks measuring more advanced theory of mind (second-order false belief tasks and understanding of sources of knowledge) and additional executive function measure of planning (additional rules in the set-shifting task and addition of the Tower of London planning task).

Before looking at the changes in children's performance over time, I consider the inter-task reliability for theory of mind and for executive functions in Hughes' study. It would be helpful to know if all of the tasks which were supposed to measure the same constructs were in fact correlated at time *a*. Hughes found moderately strong correlations between executive function measures, until age and verbal ability were taken into account. When these were factored out, there remained no significant relation between any of the executive function measures. This suggests that they were measuring independent types of competence in the children (*e.g.*, motor inhibition, cognitive inhibition, and working memory). As for the measures of theory of mind, all measures of first-order false belief understanding were significantly correlated at time *a*, even when age and verbal ability were taken into account. Performance at time *b* on classic and advanced false belief tasks was not correlated when age and verbal ability were factored out. However, there was a significant correlation between the advanced tasks (between second-order false belief prediction and understanding informational access) even when age and verbal ability were partialled out. This suggests that although classic measures of theory of mind, such as the Sally/ Anne test and the Candy test, target the same cognitive mechanisms, there is something distinct about higher-level theory of mind tasks (second-order false belief and information access tasks).

Hughes (1998) employed cross-lag correlations to determine the reciprocal relations between measure of theory of mind and measures of executive functions. She found that performance of the preschoolers on theory of mind measures was predicted by earlier performance on executive functions tasks, but the reverse was not true. In other words, children's earlier theory of mind did not seem to play a role in developing context appropriate behaviors, whereas those behaviors were a factor in developing theory of mind. This was the beginning of an approach to the question of the developmental relation between theory of mind and executive functions, which moved beyond the classic cross-sectional methodology. Another study using the same longitudinal (test-re-test) methodology supported the conclusion that executive functions predict later theory of mind, but not vice versa (Carlson, Mandell & Williams, 2004). In this replication 81 young preschool children were tested and then retested at an interval of 15 months. Their executive functions scores in the first session predicted later theory of mind scores, but not vice versa. These longitudinal studies go beyond cross-sectional correlational studies to suggest that individuals' performance on a later test is predicted by their performance on an earlier test. It remains that before I can draw conclusions as to the (causal) developmental relation between theory of mind and executive functions, we would need stronger evidence showing that *one develops before the other within individuals*. Accessing this sort of developmental picture necessitates a method of frequent sampling throughout a period of change, coupled with a single case approach.

**I.VI.2. Microgenetic studies of theory of mind development**

Flynn, O'Malley and Wood (2004) undertook to follow 21 children throughout a period of change in their mental state reasoning, rather than using a simpler test-retest design as in the Hughes study. This type of microgenetic method, focusing on individual paths of change, is exactly what will allow us to ultimately shed light on the question of causal dependence between theory of mind and executive functions. Indeed, showing a correlation, whether cross-sectionally or through time, is not the strongest evidence for a functional relation between two sets of competence. Flynn and colleagues selected children on the basis that they failed most of the tests of theory of mind at the initial testing phase, and continued to test them once every four weeks for six months on two theory of mind (unexpected events) tasks and two executive function tasks. Here, children were indeed followed throughout a period of change, but the analyses of individual development as they compared to the group averages were quite revealing. First, I discuss children's development in theory of mind. The authors revealed that although a group average at each phase would have shown steady improvement in children's mental state understanding, there were in fact several profiles. Eight children performed consistently (either failed all test questions, or passed all test questions) and four of children consistently improved, while nine other children began to pass the test questions, only to fail again at a later point in time. These microgenetic data show that children's performance during a sensitive period of developmental change is likely to progress non-monotonically. The authors did not have a way of quantifying change within single case studies. Still, they showed individual developmental paths that could not be predicted based on group averages.

Now classifying children on the basis of whether theory of mind improved first or executive functions improved first, Flynn, O'Malley and Wood (2004) showed that the majority of children showed “good” executive skills before they showed “good” theory of mind (though there were a few exceptions). Again, setting forth these individual profiles, in order to characterize a sequence of development, was only possible because of the longitudinal, microgenetic nature of the study. Still, we must have a way of quantifying variability between individuals as well as instability within individuals, if these are the developmental phenomena we face.

**I.VII. New directions: Cross-sectional group and longitudinal single case approaches to variability in executive functions and theory of mind development**

I have reviewed classic literature on children’s understanding of mental states, and I have shown that their ability to demonstrate knowledge of mental states might be causally related to executive factors. In the collection of empirical work presented in the next two chapters I employed cross-sectional group designs and longitudinal case studies to ask about the relation between theory of mind development and executive function development. What do individual’s profiles of development look like, for theory of mind and for executive functions? Are profiles of development similar within theory of mind and executive function tasks across children? Are sequences of development across tasks similar across children?

Such developmental questions can lead to a dynamic picture of how developmental change occurs, taking into account inter- as well as intra-individual variability. In Chapter II, I describe a new measure for executive functions susceptible to enter into theory of mind reasoning. Using a cross-sectional group approach, I measured a type of executive functions requiring the child to inhibit a strong, but wrong, response. The correlation between this performance on this new measure and performance on classic theory of mind measures is discussed. In Chapter III, I describe a new method for statistically analyzing the development of theory of mind on a single case basis. I analyze performance within and across theory of mind tasks and our new executive function tasks within individuals, over a period of several months. Individual developmental curves are produced to illustrate a pattern of change in theory of mind reasoning that displays a great deal of inter- and intra-individual variability. In Chapter IV, a picture of developing theory of mind emerges and implications of our findings from cross-sectional group data and longitudinal single case data are discussed.

# Chapter II

## II.I. Executive functions in development: Definitions and Import

Inhibitory processing and executive functioning more generally are central theoretical constructs in many discussions of cognitive development. Although there remain some questions about if and how these abilities can be fractioned (MacLeod et al., 2003; Diamond, Kirkham, & Amso, 2002; Friedman & Miyake, 2004; Senn, Espy & Kaufmann, 2004; Luciana, Conklin, Hooper & Yarger, 2005), what is clear is that children's ability to control their actions in a context appropriate fashion normally increases during the preschool years. Most executive function research focuses on inhibition, working memory, and planning capacities which reflect individuals' highly regulated information processing. This area of research has implications not only in normal development, but also in abnormal development. Underdeveloped executive functions can lead to behavioral deficits (e.g., impulsivity, stereotypicality, and disorganization), which are often observed in atypical development (e.g., Tourette's syndrome, ADD, OCD). Children with atypical executive functions often have difficulty acquiring new skills, as well as applying existing skills, putting them on a learning disabled path from the beginning of their school careers (Meltzer, Katzir-Cohen, Miller & Roditi, 2001; Denckla & Reader, 1993; Mahone, Koth, Cutting, Singer & Denckla, 2001).

## II.II. Executive functions and core domains of knowledge

Because of the broad-reaching implications of executive control, I think it important to ensure that a broad range of tasks are at the experimenter's and the clinician's disposal. Using diverse batteries, we can further elucidate the relation between the various executive functions themselves, as well as between executive functions and other cognitive capacities. Inhibitory processing has become a “usual suspect” to explain changes in various domains of cognitive development (see for example in the domain of *naïve physics* Hood, Wilson & Dyson, 2006; in the domain of *naïve mathematics* Espy et al., 2004; and in the domain of *naïve psychology* Leslie & Thaiss, 1992).

**II.II.1 Executive functions’ role in theory of mind development**

Among these core domains of knowledge, a great deal of work has been done to understand how naïve psychology develops. Understanding that people are agents with (unobservable) mental states, and that these mental states predict (observable) behavior requires one to apply a sophisticated body of principles, sometimes called a theory of mind. Classically theory of mind development is assessed by telling the child a short story with the aid of toys or pictures. At the end of the story, children are required to predict what will happen next, based upon the character’s beliefs and feelings (rather than their own). For example, in the ‘false belief – unexpected transfer’ task, children see a doll named Sally put her ball in a basket while her friend Anne watches. Then, after Sally leaves the room, Anne transfers the ball to another location, say, a box. Given that Sally did not witness this unexpected transfer, children are asked to predict *where Sally thinks the ball is*. If they answer that Sally thinks the ball is still in the box, then they are considered to have passed the test of theory of mind. Predicting that Sally has a false belief about the location of the ball demonstrates the ability to attribute beliefs to Sally on the basis of her informational access, even when Sally has not seen everything the child has seen. Since the inception of the false belief paradigm, there have been many studies of children’s developing theory of mind, both in normal and disabled populations (Baron-Cohen, Leslie & Frith, 1985; Hogrefe, Wimmer & Perner, 1986; Perner, Leekam &

Wimmer, 1987; Leslie & Thaiss, 1992; Clements & Perner, 1994; Roth & Leslie, 1998; Hughes, 1998; Moore, Pure & Furrow, 1990; Leslie & Polizzi, 1998; Friedman & Leslie, 2004; for a review, see Wellman, Cross and Watson, 2001). Three-year-olds typically fail to predict where Sally thinks the ball is. For a long time many authors thought that because three-year-olds cannot attribute false beliefs, they lack a theory of mind. However, it now appears that even infants appreciate a set of principles resembling mature theory of mind (Onishi & Baillargeon, 2005; Surian, Caldi & Sperber, 2007; Southgate, Senju & Csibra, 2007). Infants as young as 15 months look longer at outcomes that are not predicted by theory of mind. For example, babies are surprised when a character searches for an object in the correct location, after the object has been moved unbeknownst to the character. Indeed, according to a coherent and mature theory of mind, the character shouldn't know where the object is, and therefore shouldn't search for it in the correct location. Thus, the challenge to our field now is to explain why three-year-olds have trouble demonstrating their theory of mind within certain task constraints, while infants can demonstrate their theory of mind within other task constraints. It has been suggested that some performance factors linked to the classic verbal paradigm used with preschoolers are in fact shielding competence which has been there since infancy. Many researchers now believe that three-year-olds' performance can be explained by looking at the role of executive functions, and inhibition in particular, in children's development (Russell, Mauthner, Sharpe, & Tidswell, 1991; Leslie & Polizzi, 1998).

### II.II.2 Correlations between executive functions and theory of mind

Investigating the developmental relation between executive functions and theory of mind has thus become an important topic. Efforts over the past decade have increasingly focused on comparing children's performance on theory of mind tasks to their performance on executive function tasks (Flynn, 2006; Carlson, Mandell & Williams, 2004; Carlson & Moses, 2001; Perner, Lang & Kloo, 2002; Russell, Mauthner, Sharpe, & Tidswell, 1991; Perner & Lang, 1999; Hughes, 1998; Flynn, O'Malley & Wood, 2004; Carlson, Moses, & Claxton, 2004). In most of the correlational studies, researchers have found small to moderate positive correlations between performances on these tasks, but also frequently correlations between executive function and theory of mind scores failed to reach significance. For example, Carlson, Moses & Breton (2002) found that when age and IQ were factored out, the appearance-reality test of theory of mind did not correlate with their executive function measures ("conflict" and "delay" batteries). Also, Perner, Lang and Kloo (2002) found non-significant, at times negative, correlations between False Belief- location scores and other inhibition tasks (Card Sorting, Go/ No-Go) and when age, verbal ability and control question performance were partialled out.

Interpreting these variable results is made difficult by diverging procedures. First, the factors researchers chose to partial out of their analyses were different for each of the studies (Carlson & Moses, 2001: age, gender and verbal ability; Carlson, Moses & Breton, 2002: age and IQ; Perner, Lang & Kloo, 2002: age, verbal ability and control question scores). Second, the questions used to assess theory of mind understanding, although very similar from an adult's perspective, may lead to various levels of difficulty for children who are chosen precisely because they are in a critical time-window of development for the ability under investigation. This means that the slightest modification from one procedure to the next may induce noise. Notably, there are serious differences in the way theory of mind understanding was assessed for the false belief – unexpected transfer task. Perner, Lang and Kloo (2002) asked the children where Sally will first go to get her ball, Carlson and Moses (2001) asked the children where Sally thinks her ball is, and Carlson, Moses and Breton (2002) asked the children where Sally thinks her ball is in one version, and where Sally will first look for her ball, in another version. This particular manipulation, the addition of the word 'first' to the critical question, has been shown to affect children's performance on the false belief – unexpected transfer task (making it easier to pass), and therefore renders a comparison of these correlational findings more difficult (Siegal & Beattie, 1991; Surian & Leslie, 1999; see also Leslie, German & Polizzi 2005, for further evidence).

**II.III. Gauging executive functions in preschoolers**

These findings, taken together with evidence that executive functions enter into a wide variety of skills, suggest to us that research in many domains of cognitive development could benefit from a better portrait of executive functions. The current theoretical question regarding executive functions ("are they one or are they many?") can be addressed in several ways, starting first and foremost with a well constituted battery of tasks. It seems that there is currently a gap in the battery of measures for executive function in preschoolers. Common tasks used to gauge inhibitory capacity in preschoolers require children to apply a rule across the same stimuli on every trial, and these rules give specific instructions as to the possible stimulus-response pairings which are available to the child. Thus, current tests of inhibitory processing repeatedly present the same stimuli or require the same responses across trials, or both, testing the child's ability to associate (or overcome) the specific stimulus-response pairing. We do not have a test of inhibitory processing that varies the stimuli on every trial, and requires the child to use an abstract rule-response pairing to find the correct response given the novel stimulus. Because our goal is to add to the measurement tools used to assess executive functions, and inhibitory processing in particular, let us briefly review three classic tasks that are used to measure preschoolers' executive function.

**II.III.1. The Day/ Night task**

In the standard version of this task, children are shown two different types of cards, over a series of trials (Gerstadt, Hong & Diamond, 1994). During the initial training phase, children are instructed to say 'Day' when they see a picture of the moon on a black card, and 'Night' when they see a picture of the sun on a white card. After training on these two rules, children are tested over a series of 16 trials, with little or no feedback about the answers that are given or reminders about the rules. Children 4 ½ - years-old make errors on approximately 20% of the trials, while 4-year-olds make errors on approximately 50% of the trials. Their performance can be improved by changing the semantic relation between the card and the word children are instructed to say. That is, if the rule is to say 'Dog' when a moon card is shown, and 'Pig' when the sun card is shown, both 4- and 4 ½-year-olds perform at ceiling.(2)

**II.III.2. The Dimensional Change Card Sort task**

In the standard version of this task, children are shown two target cards, which comport two dimensions (shape and color, for example). The target cards are chosen to represent one value of each dimension. The cards may be, for example, a yellow dog and a red car. Test cards are a set of all possible combinations between the dimensions (yellow dogs and cars, red dogs and cars). Initially, children are told to sort the cards, following the two targets, according to one rule (say, color). After this phase of sorting, they are told to play a new game, in which the rule is to sort according to the second dimension (say, shape). If preschool children initially sorted yellow cards into the 'yellow dog' box, and red cards into the 'red car' box, on the second phase, they have difficulty sorting all dogs and only dogs into the previously 'yellow' box and all cars and only cars into the previously 'red' box. This perseveration on the initial rule occurs even though, when interrogated verbally, children demonstrate accurate knowledge of the rule at hand. In the DCCS, approximately 90% of three-year-old children fail to switch rules between the two phases of this game, regardless of which rule they start with (Zelazo, Mueller, Frye, & Marcovitch, 2003; Zelazo, Frye & Rapus, 1996).

### II.III.3. The Whisper task

The whisper task is probably the most abstract task that existed until now (Kochanska et al. 1996, Sabbagh et al. 2006). Children see a picture of a popular character and their urge is to shout out the name. However, they are instructed to whisper the names instead. Children must inhibit their urge to shout. The whisper task is an abstract task in the sense that the rule must be generalized across varying stimuli. A different popular character is presented to the child on each trial. However, it is not an abstract task in the sense that the rule indicates the correct course of action in the presence of a given stimulus. In whisper part of the response program is given in the rule (i.e., whisper the name rather than shouting it). Thus, the rule can be consulted to resolve the response conflict. Indeed it is the difficult part of the response program (whisper, don't shout) that is given in the rule.

### II.IV. Adding a test of abstract executive functions to existing measures

The rules in both the Day/ Night task and the DCCS task rely on perceptual pairings between stimulus and response. Although acquiring a mental representation of these pairings, and controlling their pull on our actions, represents an important part of inhibitory processing, there must be other types of rules, in other contexts, which enter into inhibitory computations. I am thinking in particular of the types of rules that play a role in theory of mind reasoning, i.e., abstract rules. There is nothing in the perceptual environment which informs our understanding of people's beliefs and desires, because people's beliefs and desires remain invisible and intangible, concepts of mental states do not follow a stimulus-response mapping. A new abstract task could be abstract in both senses I described above: The stimuli could vary on every trial *and* the rule would not give information about how to resolve response conflict. The motor plan which should be elaborated in response to a trial of abstract inhibition should be informed by a conceptual reasoning mechanism which selects the correct target. The selection of a response between two conflicting response options cannot be "read off" of the rule itself.

Our goals in the present study were twofold: first, I wanted to develop a new test of abstract inhibitory processing (given an abstract rule, not anchored in perceptual cues or motor programs, with stimulus and response changing on every trial) because such a measure seemed to be lacking in the currently available tests of executive functions in preschoolers. Putting more measures of executive functions at the researchers' and the clinicians' fingertips can help survey the frontier of this somewhat vague theoretical construct. Second, I proposed to compare performance on this new abstract inhibitory processing task with performance on classic theory of mind tasks.

## II.V. Experiment 1

### II.V.1. Method

#### II.V.1.1. Participants

50 children were recruited from area preschools (mean age 59 months, standard deviation 6.5 months, 28 girls, 22 boys) with consent from their teachers and parents. (3) The same female experimenter administered every individual session during school hours in a quiet area of the school.

### II.V.1.2. Materials

An 8x11 inch binder full of laminated pages was placed on the table in front of the child. Each page of the binder showed two pictures side by side, separated by a line. The pictures were approximately 4 inches in diameter. Some presented semantically opposed pairs of pictures (e.g., open and closed doors), while others represented variations along a relevant dimension (e.g., swimming and flying ducks).

### II.V.1.3. Procedure

The child was asked to play a game wherein the experimenter would say a word, and then show some pictures. On the child's turn, they were asked to point as a picture, any picture they wanted. The game progressed for 14 trials, where the experimenter turned the page, and showed two new pictures, naming one of them and then letting the child point at any picture. I recorded the first picture the child pointed to. If the child pointed to the picture denoted by the experimenter's spoken word, I called this choice "congruent". If the child pointed to the other picture, beside the one denoted by the experimenter's spoken word, I called this choice "incongruent". I wanted to see what children's baseline response would be, in the absence of instructions as a starting point in the construction of this new measure of inhibitory processing.

### II.V.2. Results

Each child received a score out of 14 representing the total number of times they chose to point to the incongruent picture. First I asked whether gender was a significant factor in children's performance. A one-way ANOVA revealed that gender had no effect on children's pointing choices ($F(1, 48) < 1$). For subsequent analyses I disregarded the effect of gender. In a one-sample test, I compared the mean number of times children chose to point to the incongruent picture to a theoretical mean of 7, which would represent no bias on the children's part. Children showed a significant pointing bias ($t(49) = 4.537$, $p < .001$, two tailed). In the absence of instruction, these preschoolers preferred to point to the congruent picture (mean number of incongruent points out of 14 = 4.1, standard deviation = 4.5). Children's preference did not seem to be related to their age ($r$ (50) =.137, $p = .341$). Older children did not point any more or less to the incongruent picture than younger children.

**II.IV.3. Discussion**

Children showed a strong matching bias on our baseline test of pointing in the context of the picture game. The preschoolers chose to point to the congruent picture more often than the incongruent picture. Furthermore, it seems that the baseline matching tendency was the same for all of the preschoolers, regardless of their age. Performance in this baseline condition confirms preschoolers' overall high capacity for matching-to-sample, where the sample is an auditory stimulus. This capacity is taken for granted in many clinical assessment tools, such as the Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981), and is often used in research with developmentally disabled individuals and nonhuman animals (Dickson et al., 1996; Dube, McIlvane & Green, 1992). Here in our baseline condition I confirmed that matching comes naturally to preschoolers.

I could now use this natural tendency against them just by asking them to work against their matching bias. In other words, I was able to define an "easy" rule (congruent) and a "difficult" rule (incongruent) for the subsequent test of executive function. The easy rule would require the children to point to the word the experimenter said, and the difficult rule would require them to point to the other picture.

**II.VI. Experiment 2**

**II.VI.1. Method**

**II.VI.1.1. Participants**

244 children were recruited from area preschools with consent from their teachers and parents (115 boys and 129 girls). They were divided into three age groups. There were 67 three-year-olds (mean age = 43 months, standard deviation = 3 months), 100 four-year-olds (mean age = 54 months, standard deviation = 3.2 months), and 77 five-year-olds (mean age = 64 months, standard deviation = 4 months). The same female experimenter administered every individual session during school hours in a quiet area of the school.

**II.VI.1.2. Materials**

I used the same binder as in Experiment 1, containing 14 laminated pages with two pictures on each page. 191 children played the game from the front to the back of the book, and 53 children played the game from the back to the front of the book.

**II.VI.1.3. Procedure**

Each child played the game in two blocks. There were two rules for this game, a congruent rule and an incongruent rule (see below for descriptions). Children were randomly assigned to one of four groups, defined by which rule was learned in the first block, and whether there was a rule switch for the second block. The groups were: congruent-congruent, congruent-incongruent, incongruent-incongruent, incongruent-congruent.

When the game began, the child saw the first page of the binder with two pictures on it. The experimenter explained that they would play a game. In all blocks, the first page served as a training page. If the congruent rule was first, the experimenter said "We are going to play a game. In this game, I am going to say a word. Then, on your turn, you just point to the picture I said. Ready?" And then the experimenter said the first word, followed by the reminder: "Can you point to the best picture?" If the child pointed correctly on this training page, then they were told they did a great job and they were ready to play the game for real. If they pointed to the wrong picture, then the experimenter reminded them of the rule. They were given a second chance to point to the congruent picture. If they pointed to the correct picture, they were told they did a great job and they were ready to play the game for real. If they failed a second time to point to the correct picture, the experimenter showed the child where they should point (e.g., "You see? If I say OPEN, then you point to this picture, with the OPEN door. Got it?"). The game then proceeded. Each trial consisted in the experimenter giving a rule reminder (e.g., "pointing to what I say….") then saying a word. The first picture the child pointed to always counted as the answer.

If the incongruent rule was first, the experimenter said "We are going to play a silly game. On my turn, I'm going to say a word. Then it's your turn. On your turn, your job is to point to a picture on my page. But guess what? Don't point to the one I say. You point to the other picture. OK? Whatever you do, don't point to the picture I say, you have to show me the other one! Ready?" And then, the experimenter said the first word, followed by the reminder: "Now can you show me the other one?". If the child pointed correctly on this training page, then they were told they did a great job and they were ready to play the game for real. If they pointed to the wrong picture, then the experimenter reminded them of the rule. They were given a second chance to point to the incongruent picture. If they pointed to the correct picture, they were told they did a great job and they were ready to play the game for real. If they failed a second time to point to the incongruent picture, the experimenter demonstrated where they should point, reminding them to "point to the other picture, NOT the one I say!". Then the experimenter announced they would start the game for real. Each trial consisted in the experimenter giving a rule reminder (e.g., "pointing to the other picture, not the one I say….") then saying a word. The first picture the child pointed to always counted as the answer.

The first block using the incongruent rule tested the preschoolers' ability to inhibit a prepotent response within the trial (*intra-trial inhibition*), for example pointing to the flying duck upon hearing the word "swimming". The second block represented a test of children's ability to inhibit a prepotent response between trials for half of the children, those who had to effectuate a rule-switch for the second block. For those children, it was necessary to inhibit a previously relevant rule (*inter-trial inhibition*). Some children received the congruent-incongruent condition, and were thus asked to use both intra- and inter-trial inhibition at the same time in the second block. In other words, they were required to inhibit the previously relevant "congruent" rule, while also avoiding a prepotent matching bias.

There were six test trials in the first block. At the end of those trials, the experimenter said the child had done a good job, and now they were going to play the next part of the game. Block two began, and training was repeated with the appropriate rule. There were also six test trials in block two. At the end of both blocks, children were rewarded with a sticker and a certificate of appreciation. If any child pointed more than twice before hearing the test word, their data was excluded from our analyses because they were not following the basic turn-taking structure of our game. Five children were excluded for failure to cooperate, two were excluded due to experimenter error, and one was excluded due to distracting circumstances during the test session.

**II.VI.2. Results**

Each child received two scores out of six, one for the first block and one for the second block. I also counted the number of training trials each child required in each block. If they pointed correctly on the first try, the score was one. If the child pointed to the wrong picture on the first try, then to the correct picture on the second try during training, they received a score of two. If they failed to point correctly after two tries, then they received a score of three.

**Age, Gender and Order effects**

Data were entered into an Age (3) x Block (2) ANOVA. Overall, age had a significant effect on performance in our pointing game. Training and test performance in the first block improved significantly as children got older ($F(2, 238) = 5.823$, $p < .003$; and $F(2, 238) = 6.347$, $p < .002$, respectively). Also, training and test performance improved significantly with age in the second block of the game ($F(2, 238) = 11.056$, $p < .001$; and $F(2, 238) = 8.929$, $p < .001$). As they got older, children required less training, and answered more trials correctly.

I next asked whether gender was a factor in our pointing game. A multivariate ANOVA with Gender (2) and Block (2) as factors showed that gender did not significantly affect preschoolers' performance in any of the training or in any of the test blocks (block one training: $F(1, 240) = 2.778$, $p = .097$; block one test: $F(1, 240) = .125$, $p = .724$; block two training: $F(1, 240) = .901$, $p = .343$; block two test: $F(1, 240) = .031$, $p = .861$).

Next, data were entered into an Order (2) x Block (2) ANOVA to test whether the order of presentation of our pages affected children's training or test performance in any of the blocks. I found only one significant effect of order. There was a significant effect of order on block 1 training (*F*(1, 240) = 5.004, *p* = .026). Children who saw the pages in reverse order needed less training (mean training trials = 1.19, standard deviation = .483) than children who saw the book in the original order (mean training trials = 1.40, standard deviation = .673). However, the means for both groups remained under 2, showing that on average, whether children saw order 1 or order 2, they were able to correctly learn the rules during training, either on the first or second try. Furthermore, because there was no effect of order on the number of correct answers during the test phase, I collapsed the two orders for subsequent analyses.

**Effect of the rule in the first block**

I next asked whether Rule (congruent x ingcongruent) was a significant factor in children's performance. When I looked at performance in the first block only, I found that there was a significant difference in difficulty between the two rules both for the training scores and the test scores (*F*(1, 232) = 20.018, *p* < .001; *F*(1, 232) = 61.778, *p* < .001, respectively). It was easier to learn the congruent rule (mean training trials = 1.18, standard deviation = .464) than the incongruent rule (mean training trials = 1.53, standard deviation = .741). Also, as predicted, pointing during the test phase according to the congruent rule was easier for the children (mean = 5.74, standard deviation = .557) than the incongruent rule (mean = 4.1, standard deviation = 2.316; see table 1).

Table 1.

First block performance: Mean scores out of 6 test trials, by age and rule.

| | Congruent | Incongruent |
|---|---|---|
| Three-year-olds | 5.48 | 3.15 |
| Four-year-olds | 5.78 | 4.20 |
| Five-year-olds | 5.90 | 4.82 |

**Interaction effect between age and rule**

Data were entered into an ANOVA with Age (3) and Rule (2) as factors. There was no significant interaction between the age of the children and the rule they learned (congruent vs. incongruent) when considering their performance in the training and test phases of the first block ($F(2, 232) = 2.914$, $p < .056$; $F(2, 232) = 2.467$, $p < .087$, respectively). Three-, four-, and five-year-olds appeared to find the congruent and incongruent rules equally challenging. When I removed four-year-olds from the analysis, and compared three-year-olds to five-year-olds, I found a significant interaction between age and rule in the test phase of the first block ($F(1, 140) = 6.147$, $p = .014$). Both three-year-olds and five-year-olds found the congruent rule easy (mean three-year-olds = 5.48, sd = .712, mean five-year-olds = 5.9, sd = .307). However, three-year-olds performed at chance on the incongruent rule (mean = 3.15, sd = 2.36) while five-year-olds performed better (mean = 4.82, sd = 1.78).

**Effect of switching rules for the second block**

Next I asked what the cost of switching rules was for children's performance in the second block of trials by entering the data into an ANOVA with Switch (2) as a factor. I compared children's performance when they switched rules in the second block, to performance when they didn't switch rules in the second block. I found a significant switch cost for training in the second block ($F(1, 238) = 12.289$, $p = .001$). Children required more trials to learn a new rule in the second block (mean = 1.5 trials for training, standard deviation = .79) than to demonstrate their understanding of a rule they had used before (mean = 1.2 trials for training, standard deviation = .53). Children's pointing accuracy in the test trials of the second block was also affected by a rule switch ($F(1, 238) = 13.644$, $p < .001$). Children who switched rules in the second block had considerably more difficulty in the second block of the game, irrespective of the rule they were required to apply (mean for switch group = 4.17, standard deviation = 2.167), than children who continued to use the same rule in the second block (mean for no switch group = 5.08, standard deviation = 1.783; see table 2).

**Interaction effects for age and switching rules**

Furthermore, when data were entered into an ANOVA with Age (3) and Switch (2) as factors, there was no interaction effect between the age and rule switching on preschoolers' performance in the training or test phases of the second block of our game ($F(1, 238) = 2.568$, $p = .079$, and $F(1,238) = 1.098$, $p = .335$, respectively). However, I next removed four-year-olds from the analysis, and compared three-year-olds' and five-year-olds' performance in training using an ANOVA with Age (2) and Switch (2). This revealed an interaction effect between the age group and the rule switch ($F(1, 140) = 4.987$, $p = .027$). Namely, younger children had more trouble demonstrating their understanding of a new rule during training than did older children. There was no such interaction effect on performance during the test items.

Table 2.

Second block performance: Mean scores out of 6 test trials, by age, rule and switch.

| | After no rule switch | | After a rule switch | |
|---|---|---|---|---|
| | Congruent | Incongruent | Congruent | Incongruent |
| Three-year-olds | 5.50 | 3.38 | 4.83 | 1.33 |
| Four-year-olds | 5.82 | 4.50 | 5.12 | 2.91 |
| Five-year-olds | 5.94 | 4.95 | 5.28 | 4.81 |

**Three-year-olds' and five-year-olds' ability to switch rules**

Because I found an age effect when comparing three-year-olds to five-year-olds, but not when four-year-olds were included in the analyses, I continued to compare the youngest and oldest children's performance in the second block. I found that although there was no interaction effect between age and switching on performance during test items for both rules considered together, it appeared to be significantly more difficult for the three-year-olds to switch to the harder (incongruent) rule than it was for the five-year-olds ($t(34) = -6.101$, $p < .001$). It was equally difficult for the younger and older children to switch to the easier rule ($t(34) = -.838$, $p = .408$). In sum, when comparing the youngest and oldest preschoolers, I found that there was an effect of age on training in the switch condition, and there was an effect of age on the test items of the switch condition only when children switched from the easier to the harder rule.

**II.VI.3. Discussion**

I developed a new test of inhibitory capacity in preschoolers, and I showed that I could use children's natural matching tendency against them to create a difficult rule. The congruent rule was equally easy for the younger and older children. The incongruent rule was harder for three-year-olds than it was for five-year-olds. Children had more trouble switching rules than continuing to use the same rule. I found an age difference in children's switching performance to the difficult rule, but not for switching to an easy rule. The most difficult condition in our new test was when children were asked to switch rules in the second block and apply the difficult, incongruent rule. Thus, children had the most difficulty in our task when the executive demands were high, i.e., when they had to inhibit their matching tendency to apply the incongruent rule, or when they had to switch rules between blocks and inhibit the previously relevant rule.

**II.VII. Experiment 3**

Part of our aim was to devise a new test of inhibitory processing which might tap the same abilities as Theory of Mind development. Therefore, I compared children's performance on our new test of inhibitory processing to children's performance in classic Theory of Mind tests.

**II.VII.1. Method**

**II.VII.1.1. Participants**

104 preschoolers (mean age = 51 months, standard deviation = 7 months) were recruited from area schools with prior written consent from their parents.

**II.VII.1.2. Materials**

I used the same binder as in experiments 1 and 2.

**II.VII.1.3. Procedure**

They were seen individually in a quiet area of their preschool by the same female. Children were rewarded with a sticker and a certificate of appreciation at the end of their session. The testing session lasted about 30 minutes when children received all of the tasks, and it lasted about 15 minutes for children who received all but the vocabulary test. Three children were not given all of the games because the experimenter determined they were not cooperating. Five children were excluded because we ran out of time for the test battery. Except for the vocabulary test, only children who completed all of the games were included in our analyses. (All of the children who started the vocabulary test finished it.)

I first gave children the Peabody Picture Vocabulary Test- Revised to determine their verbal mental age (Dunn & Dunn, 1981). A subset of children did not receive the PPVT because of time constraints in their schedule. All children received two classic false belief tasks and our new test of inhibitory processing. The false belief tasks were an unexpected transfer and an unexpected contents task (based on Baron-Cohen, Leslie & Frith, 1986, and Hogrefe, Wimmer & Perner, 1986). In the false belief - unexpected transfer task children saw a doll named Sally put her ball in a basket while her friend Anne watched. Then, after Sally left the room, Anne transferred the ball to another location, say, a box. Given that Sally did not witness this unexpected transfer, children were asked to predict where Sally thinks the ball is. If they answered that Sally thinks the ball is still in the box, then they passed the test of false belief attribution.

In the false belief - unexpected contents task children were shown a familiar container, say, a candy tube. The child was asked to state what they thought was inside (candy). Then, the experimenter opened the tube to reveal an unexpected contents, say, a pencil. Children were then asked to report their own prior false belief, that is, what they first thought was in the container (candy). Finally, children were asked to report what someone else would think was in the container (candy).

Children were given the difficult condition of our new inhibitory processing task, namely, the congruent- incongruent condition, where they had to switch rules from the easier to the harder rule. In that game, children first were trained with the congruent "point to what I say" rule, then played the game for six trials with six different pairs of pictures. At the end of those six pairs, the rule was switched and children were trained with the incongruent "point to the other picture" rule, then played the game for six trials with six new pairs of pictures.

In the false belief tasks, children received one point for each correct answer they gave to a belief attribution question. That meant that they could receive exactly one point in the false belief- unexpected transfer game and they could receive two points in the false belief- unexpected contents game. If a child failed control (memory) questions after being given a second chance, they were considered to have failed the test question no matter what answer they gave. They could receive up to six points in the first (congruent) block of the inhibition task, and they could receive up to six points in the second (incongruent) block of that task. PPVT scores represented the age equivalent to their vocabulary in months.

**II.VII.2. Results**

**False belief – unexpected transfer and unexpected contents**

Not surprisingly, children's performance on the two false belief tasks was moderately correlated ($r(91) = .612$, $p < .001$; see table 3). Both the false belief – unexpected transfer and the false belief – unexpected contents tasks were positively correlated with age ($r(91) = .309$, $p = .003$, and $r(91) = .381$, $p < .001$, respectively).

**Inhibition task – first congruent block**

Next I looked at children's performance on the first block of our inhibition game when children had to apply the easy congruent rule. I only found one significant correlation involving performance on the first congruent block of our game. Performance on the first (congruent) block was positively correlated with age, ($r(91) = .269$, $p = .01$). Although mean performance was high across ages on this task (mean = 5.6/ 6), the standard deviation was also high (sd = .94).

**Inhibition task - second incongruent block**

I then looked at preschoolers' performance on the second block of our new inhibitory processing task under the difficult incongruent rule, when a rule switch was also required. I found a significant positive correlation between performance on the second block of our game (when children had to switch rules and apply the incongruent rule) and performance on both the false belief- unexpected transfer task ($r(91) = .304$, $p = .003$) and on the false belief- unexpected contents task ($r(91) = .364$, $p < .001$). This part of the inhibition game was also correlated with age ($r(91) = .288$, $p = .006$), and with vocabulary ($r(91) = .249$, p = .017).

Table 3. Correlation coefficients between our tasks. Partial correlations, controlling for age and vocabulary are shown in parentheses.

| | Congruent first block | Incongruent second block | False belief- unexpected transfer | False belief- unexpected contents |
|---|---|---|---|---|
| False belief- unexpected contents | .169 (.063) | .364** (.257*) | .612** (.491**) | |
| False belief- unexpected transfer | .132 (.042) | .304** (.206) | | |
| Incongruent second block | .031 (-.057) | | | |
| Congruent first block | | | | |

* $p < .05$, ** $p < .01$

**Partial correlations – factoring out age and vocabulary score**

Next, I asked whether the positive correlations between the theory of mind tasks and the inhibition tasks were attributable to other factors. Two obvious and quantifiable factors which could underlie children's performance in both the inhibition task and the theory of mind tasks were children's age and their verbal abilities – especially given that performance on parts of the inhibition task and the false belief tasks were shown to correlate with both of these factors. Therefore, I carried out a series of tests to control for age and verbal abilities. When accounting for age and PPVT scores, I found the correlation between our two measures of false belief remained strong ($r(87) = .491$, $p < .001$).

The partial correlation (controlling for age and vocabulary score) between the first block of our inhibitory measure (the congruent baseline condition) and our Theory of Mind measures were first carried out. I found no significant relationship between performance in the congruent condition and false belief- unexpected contents scores, nor between performance in the congruent condition and false belief- unexpected transfer scores ($r(87) = .063$, $p = .557$, and $r(87) = .042$, $p = .698$, respectively).

The partial correlation (controlling for age and vocabulary) between the second block of our inhibitory measure (the incongruent switch condition) and the false belief – unexpected contents task was significant, ($r(87) = .257$, $p = .015$). There was also a weak positive correlation between the second incongruent block of our inhibitory measure and the false belief – unexpected transfer task, but it fell just short of statistical significance ($r(87) = .206$, $p = .053$). I found the same pattern when age alone was partialled out. The correlation held between the second incongruent block and the false belief-unexpected contents task ($r(87) = .287$, $p = .006$), as well as between the second incongruent block and the false belief unexpected transfer task ($r(87) = .237$, $p = .025$). In other words it seems that the significant positive relationships I observed at first between these tasks were not solely due to children's increasing age and verbal abilities. There is a part of shared variance between our false belief measures and our inhibitory measures which is independent of preschoolers' age and verbal ability.

**II.VII.3 Discussion**

Our inhibition test was comprised of an easy block (congruent rule - presumably requiring few inhibitory resources) followed by a difficult block (incongruent rule - presumably requiring inhibition of the previously relevant easy congruent rule, in addition to requiring inhibition of the currently present yet irrelevant response option).

This time, performance in the first congruent block correlated with age. In experiment 1, where children were free to point wherever they wanted, and children chose to point to the congruent picture (matching bias), there was no correlation with age. I believe it is possible that adding the explicit "congruent" rule in experiment 3 overloaded children's capacity to pay attention to a rule while pairing a stimulus and a response. Apparently children were better at applying the congruent rule as they got older.

The second block, with higher inhibitory demands, but not the first block, with low inhibitory demands, of our new test of inhibitory capacities in preschoolers correlated significantly with their false belief attribution skills. This type of inhibition is presumably similar to what is required for theory of mind, if there are any executive components at all. In the second incongruent block of our inhibition game, children had to keep in mind that they were to shift away from the picture which matched the stimulus word they heard, and point to the other picture. This would apply across different stimuli, as they heard a new word and saw a new picture on every trial. Similarly, in theory of mind reasoning, contexts are always shifting and there are no perceptual cues which will allow a child to identify which belief or desire or other mental state is the correct one to attribute. They must rely on an abstract understanding of these concepts, and this understanding may require them to overcome their own current mental states in order to properly attribute mental states to others. Furthermore, not surprisingly, our two theory of mind tasks were correlated. This supports the idea that the false belief – unexpected transfer and the false belief – unexpected contents tasks measure the same underlying conceptual understanding.

## II.VIII. General Discussion

### II.VIII.1. Abstract executive functions in preschoolers

I set out to create a new measure of inhibitory processing that did not rely on children's ability to associate perceptual stimulus-response pairs. I found that performance improved significantly on our new task through the preschool years. Younger children had lower scores on our "easy" congruent rule, compared to the older preschoolers (though overall scores were high for the first congruent block of trials). It is possible our youngest three-year-olds had trouble with this task because they had trouble paying attention to the rule and the stimulus at the same time in order to select the appropriate answer. Moreover, younger children had significantly lower scores than older children on our "difficult" incongruent rule. They had trouble pointing to the incongruent picture in the first block of the game, and if they had to do this in the second block, after switching rules, their difficulty was compounded. These age trends suggest I targeted the correct age group for our new measure of abstract inhibitory processing. I also found a rule x age interaction, when I contrasted performance of three-year-olds and five-year-olds. I found that the congruent rule was equally easy for all of the preschoolers, while the incongruent rule elicited chance performance from the three-year-olds and good performance from the five-year-olds. This new measure could be used in a variety of contexts where it is desirable to assess executive capacities in preschoolers.

**II.VIII.2 Relation between performance on abstract executive functions tasks and theory of mind tasks**

After establishing baseline behaviors and constructing a new measure of inhibitory functions in a fully crossed design with ‘switch’ (switch vs. no-switch) and ‘rule’ (congruent vs. incongruent) as factors, I compared performance on this new inhibitory task with performance on theory of mind tasks. Our findings suggest that the executive processes recruited by the difficult conditions in our task (when children must apply the difficult incongruent rule after a rule switch) have something in common with those recruited in reasoning about mental states. The shared variance between the tasks remained even when I controlled for age and vocabulary level. Although a portion of variance may be shared between theory of mind and our abstract inhibition task, there remains a substantial part of variance which is not shared. Therefore, our findings along with previous research on correlations between theory of mind and executive functions continue to support the idea that there are at least some overlapping processes between theory of mind and inhibition. Further studies are needed to tease apart the nature of this relationship (Friedman, Baker, & Leslie, submitted).

Having a broad range of measures at our disposal will help determine which processes are involved in development of domain-specific capacities, such as theory of mind. This new test of abstract inhibitory capacities contributes a new way of measuring executive functions, and to our knowledge it serves as the first test to target preschoolers’ use of abstract rules.

**II.VIII.3. Refining the new measure of abstract inhibition**

There are two empirical advances offered by the opposites task. First, we have a baseline measure of children's tendency to match the word they hear and the picture they point to. The Day/Night and Whisper tasks do not have such a baseline, though the dimensional change card sort does. The baseline condition of the opposites game allows us to quantify the urge we are asking the children to go against during the more difficult incongruent trials. Indeed it is difficult to talk about inhibition without describing what is being inhibited.

Related to the baseline block is another empirical advance offered by the opposites game. It contains two measures of inhibition, one intra-trial and one inter-trial inhibitory effort. The Day/Night and Whisper tasks only offer one measure of inhibition (intra-trial inhibition). Being able to measure two kinds of inhibition with one task is efficient and allows us to compare these two types of inhibition without introducing additional factors by using different paradigms.

In opposites, the intra-trial inhibitory effort is required in the first block of the incongruent rule, where children must overcome their impulse to point to the picture that is named, pointing instead to the other one. This is analogous to the Day/Night and whisper games. However, the second block of the opposites game where children switch rules (operating inter-trial inhibition) and must inhibit a previously relevant response, has no analog in the Day/Night or Whisper games. Only the Dimensional Change Card Sort offered this rule-switching.

Our results show that intra-trial inhibition shares more variance with theory of mind reasoning than inter-trial inhibition does. In this respect, the opposites game and the Day/Night and Whisper games are equally relevant to theory of mind measures. Still, in the name of exploring executive functions and how they develop in preschool, it is desirable to have a measure of both intra- and inter-trial inhibition based upon an abstract rule, like in the opposites game.

Using this array of measures to assess children's EF can help us to better understand EF development. Any new ways of measuring EF may inform us about the processing capacities which come into play in theory of mind reasoning.

A crucial part of exploring any new measure is to relate it to existing ones. Thus, it seems important to compare this new "abstract" test of inhibition using the opposites game with other measures of inhibition that are typically used in theory of mind experiments. A further study is planned, exploring the correlations between opposites, Day/Night, the Dimensional Change Card Sort, Whisper and theory of mind. Vocabulary will also be measured. Given the sheer length of these tasks, I propose to administer them in two separate sessions on the same day, with a short break in between.

**II.VIII.4. Executive functions: Theoretical construct in progress**

Questions still remain as to how we might best delineate the executive functions. Classical views describe at least two types of executive functions: working memory and inhibition, and some add planning or problem solving (Luria, 1973). I have opted to discuss the executive processing requirements in terms of inhibition. There are indeed alternative formulations of the executive mechanisms at work. As Leslie, German and Polizzi (1995) put it: "…we could say that the true-content is more highly activated than the false-content. Or we could say that the true-content has a higher subjective probability (of being correct) than the false-content. Then instead of inhibition we could speak of lowering the activation level or decreasing subjective probability

Or we could say that attention has positive and negative polarities." Because these alternatives are equally plausible given empirical evidence, I chose to speak in terms of inhibitory processing as it is described in the ToMM-SP model.

But this does not entirely eradicate the problem of drawing up definitions. Inhibitory processing itself can be further divided into (at least) two sub-components: inhibition of competing responses ('Conflict' inhibition by versus inhibition of one response ('Delay' inhibition; Carlson & Moses, 2001). Grouping the inhibition measures under these two headings, Carlson & Moses (2001) found a stronger correlation with theory of mind tasks for the 'Conflict' battery than for the 'Delay' battery, although both correlations were statistically significant. Carlson, Moses and Breton (2002) found that children's performance on two theory of mind tasks, (appearance- reality and false belief-unexpected transfer) correlated significantly with the 'Conflict' inhibition measures, but not the 'Delay' inhibition measures. These findings show that there could be a multitude of executive functions, and they leave open the possibility that there are in fact dedicated executive functions for each type of task.

Whether executive processes are task-specific or shared between tasks, how these processes interact with other non-executive sets of abilities, and how these relationships develop over time, are central questions for future research. Indeed it is always possible that we will find one way to delineate executive functions for young children, which is not precisely commensurate with the set of executive functions we observe in adults.

## Chapter III

### III.I. Motivations for the study of variability in individual development

As we have seen, there is consensus that there are important developmental changes in preschoolers ability to pass false belief tasks (Wellman, Cross & Watson, 2001). Recently, researchers have begun to examine the processes of developmental change more closely, primarily through microgenetic studies that follow a group of children longitudinally for a period of repeated testing with the same tasks (Seigler, 1996). Such studies can give information about the shape of developmental curves over a period of important change. However, there are serious limitations in the use of group data for this purpose. First of all, development and learning are processes that take place within an individual child, and are not properties of a group mind. The use of group data is actually motivated by statistical concerns, yet it is well known that statistical findings from groups can be misleading when applied to single cases. For example, when performance over time is averaged across a group the resulting curve of mean success may rise smoothly and steadily. Nevertheless, projecting from this data to conclusions regarding individual development is ambiguous: Does performance in every single case rise smoothly over time? Or does it rise abruptly in each individual with the change point distributed smoothly across individuals over time? These two developmental patterns are dramatically different, yet it is hard with group data to disambiguate the two.

Group data is used because statistical techniques cannot be applied to single events. Here I solve this problem by adapting a technique that has recently been introduced in the study of animal learning (Gallistel, Fairhurst & Balsam, 2004). An individual subject learns over a number of trials and produces a cumulative record of their responses, so that any point on the learning curve represents the accumulation of past responses to that point. The Gallistel algorithm iteratively examines this curve starting at the origin and moving along the curve testing for changes in slope. The tests employed are inferential statistical in nature essentially examining for any given point in the cumulative curve whether responding after that point is statistically different from responding prior to that point. It is thus possible to identify for a single subject at a given level of statistical significance whether and where change points occur in the subject's response patterns across developmental time. Instead of group data which uses many individuals tested once, this procedure uses a single individual tested many times.

With this technique I examined developmental curves of single cases in performance on executive functioning and classic false belief tasks, repeatedly testing the same individual over a period of months to obtain the cumulative record for that individual. I then use inferential statistics to identify change points in that individual's record. I am therefore able to make statements regarding the course of development in single cases. By studying a number of individuals separately, I ask how frequent patterns identified in case studies are across individuals. By studying curves for executive functions and theory of mind I am able to relate the change points on each within a given individual, rather than relying on ambiguous group averages. The microgenetic method coupled with the change point algorithm can give a new picture of developmental change in theory of mind and executive functioning.

### III.I.1. What do traditional group-averages tell us about individuals?

The developmental questions we can ask are constrained by our means of empirical investigation. For instance, asking what sort of behavior or cognitive strategy is typically attained by children at a certain age would require that one has a random sample of children from which to infer typicality (or normality). In contrast, asking what part of the child's cognitive development is necessary and what part is merely a contingency would require one to look more carefully at individual differences. Are there various ways of attaining the same mature performance, or is there only one path to success? Asking what individual children may do differently, to ultimately arrive at the same result, necessarily entails a method which allows individual differences to be highlighted. Although there has been much research on what children have in common, looking at group averages and extracting normal developmental milestones from them, the field of cognitive development has given fairly little attention to variations across individual's development. Even if there is little variability across individuals, there could be variability within individuals. Is individual change monotonic and gradual, or is it abrupt? These variations can and should be an informative part of our empirical studies of change. As Flavell and Wohlwill (1969) noted: "Such flexibility in potential means could conceivably be highly adaptive for our species, given all the individual variation in its members and in their life experiences… A diversity of ontogenetic routes might thus be plausible as well as possible, and empirical probes for the existence of these several patterns might accordingly be a worthwhile research venture."

**III.I.1.1. Defining the norm in preschoolers' theory of mind**

I am interested in an ability which develops early, without schooling, and universally in normal children. The appreciation of people as mental agents, with thoughts and feelings that relate to their actions in a principled way, is sometimes called theory of mind. This naïve theory of how the mind works is increasingly expressed by children in the preschool years, through their own verbalizations and actions (Bartsch & Wellman, 1989). Do all children develop their theory of mind in the same way, or are there important individual differences? The answer to this question will inform our understanding of how change occurs, how to recognize it, and what motivates it. These are important insights for clinical work as well as for education. The methods we utilize to answer these questions for theory of mind could be an equally useful means of investigation in other domains, such as naïve mathematics and language development. I first review the cross-sectional evidence for developmental change in theory of mind, then I discuss the few longitudinal studies which have captured theory of mind development. Finally, I propose a new way of investigating developmental change under a microscope, and I ask what current theories of theory of mind development imply about individual change.

In theory of mind research, many group studies tell us that young three-year-olds fail to attribute a false belief to a character in a story where the character is absent when an object is moved. This unexpected transfer task requires the child to put aside their own knowledge about the current location of the object, say a ball, and appreciate that the other person is mistaken about where the object is. When the child must predict where the character will go to find the ball (action prediction question), or where the character thinks the ball is (belief attribution question), group data show that approximately 20% of three-year-olds correctly predict the action of or attribute a false belief to the character (see Wellman, Cross & Watson, 2001 for a meta-analysis). Group data show that when three-year-olds are asked to consider this false belief-unexpected transfer task they reply that the character will behave or think just as if the character actually had a true belief about the object's location, and that the character will look in the object's current location to retrieve it. In other words, we say that three-year-olds typically fail the false belief attribution test because <u>most</u> three-year-olds fail such a test. It is said that five-year-olds typically pass the false belief attribution task, because group data show a high rate of passing amongst five-year-olds, most averages approximating 80%. This well-established increase in performance on many different tests of theory of mind is found over the course of just two years in a lifespan of 5 or 6. The cross-sectional result has been replicated across many different environments (Sabbagh, Xu, Carlson, Moses, & Lee, 2006; Pears & Moses, 2003; Hughes, & Cutting, 1999; and see Wellman, 1998, for a discussion of how universal principles may still be subject to enculturation).

**III.I.1.2. Group-average data cannot address inter-individual variability**

Cross sectional studies are the usual way to approach the question of development, by looking at snapshots of behavior before and after change. These research designs can tell us what uniformly poor or uniformly good performance looks like. But usually performance is not uniform. For example, 30% of three-year-old may pass a task. Does that mean the children in the 30% subgroup always pass while the others always fail or that all of the children have a 30% chance of passing, or some more complex pattern? Or to take another example, 80% of four-year-olds may pass. Does that mean 80% of the children will always pass, all the children pass 80% of the time, or 50% pass all the time and the other 50% pass 60% of the time, and so on? We simply can't tell from such data. And yet for developmentalists the differences between consistent failing versus consistent passing versus consistent inconsistency are dramatic.

**III.I.1.3. Group-average data cannot address intra-individual variability**

Until now I've discussed the question of how group averages reflect inter-individual variability. What about intra-individual variability? There is some indication from group average studies in theory of mind that individual children's performance is variable. When children are tested more than once within a short time on classic false belief attribution or false belief action prediction tasks, scores for individual children tend not to be perfectly consistent. We do not observe a bimodal distribution of responses, with one group of individuals always failing the tasks, and another group always passing the tasks. For example, Siegal and Beattie (1991) found that 20% of the three-year-olds they tested failed one false belief task and passed another one. Frye, Zelazo, and Palfai (1995) found that three-year-olds gave an average of one or two correct answers out of four, for four different types of theory of mind tasks.

These results suggest that individual performance is somewhat variable: children must not all develop in an abrupt way, either finding themselves in a failing or passing category. There is good reason to believe that developmentalists will need the means to investigate both inter-individual and intra-individual variability. Our cross-sectional group methods of studying development do not allow us to explore these two types of variability. For this, the same children must be tested repeatedly throughout a period of change. We cannot know what changes are taking place in the child's cognitive structures unless we examine these changes very closely. Most of the time, children in group studies are only tested once and their performance is taken into group averages to reduce the noise and to permit statistical tests across a large number of data points. The cross-sectional approach to studying development is informative, but it does not tell us about what changes individual children undergo, since they are only tested one time. Longitudinal studies provide a means for observing children's changing cognition while it is happening. In cases where performance of an individual child cannot be captured with one data point, or in cases where performance of individual children is not represented by group averages, testing each child only once or twice will obscure the phenomenon we are most interested in studying. In these cases, the underlying behavior of individual children who are in the process of changing is so variable that a group average, based on just one or two trials per child, cannot accurately inform researchers about the process of change in the individual. Sampling individual behavior repeatedly would then be a desirable means of investigation.

### III.I.2. Individuals' profiles at center stage

#### III.I.2.1. Parallel reasoning in the animal learning literature

A parallel problem has been addressed in animal research, where learning is typically described using group-learning curves (Gallistel, Fairhurst & Balsam, 2004). The authors demonstrated that although group-learning curves typically present a model of acquisition that increases gradually and monotonically, individual animals' learning curves often show abrupt changes, and sometimes even reversals in acquisition. Thus, modeling acquisition based on group-average learning curves is misleading. Gallistel and colleagues developed an algorithm which focuses an individual animal's acquisition behavior, rather than on groups (Gallistel, Fairhurst & Balsam, 2004; Papachristos & Gallistel, 2006). According to the method they propose, quantifying change in individual mice (or rats, or pigeons, etcetera) becomes possible as a means for exploring change at the level at which it occurs, i.e., in individuals, rather than in groups.

#### III.I.2.2. Microgenetic approach to child development

Robert Siegler has been an influential proponent of studying change in a dynamic context, measuring children's performance intensively throughout a period of change (Siegler, 2007). According to the Siegler's microgenetic approach to studying development, it is important to observe the relative frequency of answers, and how the relative frequency changes through time. In this way, Siegler and colleagues have been able to show that children begin their approach to many Piagetian tasks by adopting a failing strategy. As time progresses and they are exposed to the same task repeatedly, there is a proliferation of strategies at the child's disposal. This "wave" of new insights finally leads to the child settling on a new preferred strategy for solving a problem, which is more complex and more successful than the initial strategy (see for examples Siegler & Svetina, 2002, and Siegler,1996). Without testing children repeatedly on the same set of tasks, it could never have been guessed that their progress is analogous to waves of strategic development and selection. With a cross-sectional sample, one could not have seen how individuals' strategies come and go. Thus, the microgenetic method of sampling behavior intensively throughout a period of change is very informative when it comes to the question of what development looks like in individuals.

There are a few microgenetic studies focusing on theory of mind development. Flynn, O'Malley and Wood (2004) selected 21 preschool children on the basis that they failed most of the tests of theory of mind at the initial testing phase. This maximized their chances of observing change. They continued to test these children once per month for six months on two theory-of-mind (false belief- unexpected events) tasks and two other tasks (discussed later). The authors revealed that although a group average at each phase would have shown steady improvement in children's mental state understanding, there were in fact several individual profiles. What is most striking about the data from individuals is the variability, both within and between children. There is a tendency for some children to go through a period of unstable performance, sometimes failing and sometimes passing. In Flynn, O'Malley and Wood's (2004) microgenetic study, eight children performed consistently (either failed all test questions, or passed all test questions) and four children consistently improved, while nine other children began to pass the test questions, only to fail again at a later point in time. These microgenetic data show that children's performance during a sensitive period of developmental change is likely to progress non-monotonically. Data from children in this case could have been smoothed out, as noise would be, were it not for the single case analyses. The authors did not have a way of statistically analyzing change within single cases, but they showed that the paths were not predictable based on group averages.

Flynn (2006) tested preschool children repeatedly for five months on several theory of mind tasks. Children could score one point for each answer, and received a score between 0 and 7 for each testing session. The group data from this longitudinal study showed that individuals' performance from one session to the next could vary a lot. Only 39% of the session-to-session transitions revealed the same score on the theory of mind battery, meaning that 61% of the time, children's performance changed from one month to the next. Therefore, most of the time, children's performance was unstable. Thirty-seven percent of changes from one session to the next were improvements, while 24% of changes were regressions. Also, although performance did not improve monotonically in this microgenetic study, change tended to be gradual (50% of the improvements were by 1/7 point, while only 4% were by 4/7 or 5/7 points). Again, Flynn (2006) did not have a way of statistically analyzing single case change curves.

Amsterlaw and Wellman (2006) tested 36 preschoolers twice per week for 6 or 7 seven weeks. These children received implicit feedback on the false belief scenarios they made predictions about, and they engaged in explanations regarding the characters' actions. The authors found similar intra-individual variability as Flynn and colleagues (Flynn, O'Malley & Wood, 2004; Flynn, 2006). In fact, none of the children enrolled in their study showed abrupt changes in performance, moving directly from consistently failing to consistently passing: "…all [children] proceeded through some period of intermediate success" (Amsterlaw & Wellman, 2006, p. 159). Here again, they did not have a way of statistically analyzing single case change curves.

It is increasingly clear, then, that we have a lot to gain from longitudinal microgenetic methods, which repeatedly test children throughout a period of change. Based on previous cross-sectional and microgenetic work, my emerging hypothesis is that there is a high degree of intra-individual instability together with inter-individual variability before asymptotic success is achieved. If we are to characterize the course, duration, and variety of the change processes that occur, it will be important to measure change statistically in each individual child.

### III.I.3. What do current theories predict about the shape of individual developmental profiles?

#### III.I.3.1. Two positions on theory of mind development

Are the developmental changes in theory of mind qualitative in nature, or are they quantitative? There are two major positions about the developmental nature of the change that takes place in the individual. One position says that children perform poorly on the false belief task when they are three-years-old because they do not understand the nature of belief. They cannot reason about mental states in an adult-like way because they have not yet constructed an appropriate theory (Wellman, 1988). Developmental change on this position is characterized as similar to conceptual change in science: children make observations and consider possible explanations for what they observe, testing and refining hypotheses until insight into the correct theory is achieved. Once this occurs children are able both to pass false belief tasks and to justify their answers.

The other major position is that young children are predisposed to construe the behavior of other people as arising from mental states like beliefs and desires without first having to construct theories. The basic ideas in theory of mind instead spring from evolved mechanisms of social cognition that function more like instincts than like scientific theories (Leslie, 1987). In other words, theory of mind constitutes a core domain of knowledge, much like what Gelman and colleagues have described (see for examples Gelman & Williams, 1998; Gelman, 2000). This early competence view then draws on accounts of limited processing resources to explain the failure of three-year-olds together with increasing resources to explain later success (Leslie, 2000).

**III.I.3.2. Current positions are neutral regarding rate and shape of change in individuals**

I am not concerned here with the contrasting details of these two accounts. The reason for this disregard is straightforward though in some ways a little surprising. I argue that it simply isn't clear that either of these positions makes unequivocal predictions regarding the pattern of developmental change that should be seen in an individual child. A theory of developmental change should specify the rate and type of change curve a given ability undergoes, in addition to addressing the question of the nature of cognitive processing at particular time slices. It would be natural to expect the 'theory-theory' to specify sudden abrupt insight as the underlying motor of change and therefore to predict an abrupt change pattern. However, theory-theory is free to draw upon notions like "auxiliary hypotheses" which are used to modify the child's predictions from theory on a local or temporary basis, in addition to the "paradigm shifts" in theory that underlie sudden insight, and indeed theory-theory has already done so (e.g., Gopnik, 1993). In such versions, the 'theory-theory' could accommodate almost any pattern of change. In a similar vein, the early competence view might naturally be expected to predict gradual incremental change as performance resources increase. But this kind of account is free to draw upon notions such as resource thresholds and task demands to account for abrupt changes or fluctuations in resource availability to account for unstable performance, and indeed early competence theory has already done so (Leslie, German & Polizzi, 2005). My conclusion then is that currently neither of the two main positions has addressed the question of what developmental change should look like in single case profiles seriously enough to make unequivocal predictions.

### III.I.4. Statistically analyzing individuals' developmental change

#### III.I.4.1. Two features of our new method

I wished to see if I could replicate findings from previous microgenetic studies of theory of mind, while making a number of changes to the approach other authors have used. First, I employed a new statistical technique which allows us to analyze changes in an individual's cumulative record of performance. This differs from previous methods because until now microgenetic studies of theory of mind development have needed to rely on group profiles in order to draw conclusions based on inferential statistics. Our statistical tool, borrowed from researchers in animal learning (Gallistel, Fairhurst & Balsam, 2004), makes it possible to draw conclusions about an individual's developmental curve, based on statistics. In order to use this statistical tool, though, I needed a large number of data points from every single individual. For this reason, I implemented a second change from the previous microgenetic investigations of theory of mind: I set out to test children for as long as possible, not limiting the number of sessions or the overall length of participation. Focusing on individual's cognitive performance, as opposed to group averaged cognitive performance, may be novel in cognitive developmental research, but it is a well-established means of gaining insight into neuropsychological mechanisms in brain damaged adults where the method is called "single case designs" (Shallice, 1988). Having a statistical tool which allows us to use inferential statistics on an individual child's cumulative record of performance is critical, because it allows us to give more than a description of what development actually looks like on an individual level.

### III.I.4.2. Preliminary data from single case studies

Preliminary data was collected by Rein (2005) using this approach. He gave children two classic false belief tasks per session (unexpected transfer and unexpected contents), about every two weeks. He used the Gallistel change point algorithm to reveal a high degree of inter-individual variability. While most children passed the transfer task at a higher rate earlier than the contents task, 36% of the young preschoolers showed the reverse sequence of development. This inter-individual variability was accompanied by intra-individual variability, which was quantifiable with the change point algorithm. On nine individual records performance was neither at floor nor ceiling throughout the entire testing period. That is, for nine children out of 17 tested for approximately six months, performance was characterized by a rate of passing between 20% and 80% of trials, without a statistically significant change.

## III.II. Cohort 1: Studying individual variability in developing theory of mind

One might wonder whether the inter- and intra-individual variability observed in Rein's (2005) microgenetic tests of the child's development of false belief understanding is a result of the method he used. Perhaps children, who are exposed to the same task repeatedly, as are children who participate in a microgenetic study, eventually become confused and begin to respond differently because they begin to question the purpose of the repeated questioning. To test this, I would need to include in our battery of tests one where children were expected to perform well throughout. For this reason I decided to include a true belief attribution task in our battery. Group data show that even young three-year-olds should perform uniformly well on this task. If children perform well throughout their enrollment in our microgenetic study on the true belief task, then I will be more comfortable concluding that the variability observed in false belief paradigms over time is not created by repeated testing. I could then make a more convincing argument regarding the theoretical relevance of the variability in the false belief performance of individual children over time. I also included a classic measure of false belief performance, in order to see if I could replicate previous results. And finally, I wanted to see whether I could use this new method to establish a sequence of development across abilities within individuals. In order to establish this new method as a viable one, I chose tasks whose developmental relation has been investigated cross-sectionally within groups, to see if I could confirm the sequence longitudinally within individuals. I chose the false belief unexpected transfer "look first" manipulation of this task, because it is suggested by cross-sectional literature that this task develops earlier than the classic version.

Siegal and Beattie (1991) modified the classic false belief- unexpected transfer task, with the supposition that three-year-olds might fail the classic action prediction question "where will Sally look for the ball?" because of their interpretation of the speaker's intent. The authors suggested that young children may misuse Gricean maxims to infer the speaker's intent when they hear the experimenter's question "where will Sally look for the ball?" (Grice, 1975, as cited in Siegal & Beattie, 1991). When the child attempts to interpret this question, they may assume the experimenter really means "where will Sally have to look in order to find the ball?". Indeed it has been shown that young children interpret utterance implicatures differently from adults (Noveck, 2001).

To test this explanation of why three-year-olds failed the classic belief attribution question, Siegal and Beattie (1991) used a question of action prediction which should make the speaker's intention more clear, namely, "where will Sally look *first* for the ball?". Now children must interpret the question according to the speaker's intent. Three- and four-year-olds answered 71% percent correct in the new "look first" condition, compared to 35% correct in the classic "look" condition. The effect of this minor change in the action prediction question in the false belief task attracted the attention of other researchers. Surian & Leslie (1999) replicated this finding, showing three-year-olds passed the classic "look" question only 30% of the time, but 83% passed the more explicit "look first" question. In their second experiment using a within-subjects design, they found that young three-year-olds were significantly more likely to pass the "look first" version and fail the standard task than the reverse. Based on cross-sectional data, I expected that our new method of analyzing individual's cumulative records of performance would reveal that individual children developed high rates of passing the "look first" manipulation before they developed high rates of passing the classic version.

In sum, our aim was to explore the rate and shape of acquisition in false belief understanding using a modified version of the microgenetic method with the possibility of statistics on individual's records, including a control condition (true belief task), and two test conditions (false belief "look first" task; classic false belief task). I expected that I would observe variability in individual performance in our test conditions, but not in our control condition. Further, I hoped to reproduce the sequence of development (true belief before false belief "look first" before classic false belief) within individuals that is suggested by cross-sectional group averages. This would test the credibility of our new microgenetic method.

**III.II.1. Method**

**III.II.1.1. Participants**

After obtaining permission from preschool directors, I sent recruitment materials to all children between 3;0 and 5;11 in each school. This age range was chosen to maximize our chances of observing change during the period of testing. Cross-sectional and longitudinal studies of theory of mind in preschoolers indicate that there is a shift in performance on classic false belief tasks, from consistently failing to consistently passing, around the age of four. Because this age is based on group averages, I surmised that any individual child between 3;0 and 5;11 at the time of enrollment in the study might undergo a change in performance. Fifteen children agreed to participate. For statistical reasons, it was necessary to exclude children who participated in less than five testing sessions. Some children missed sessions and were not able to make them up due to scheduling conflicts in their attendance at the school (6 children) or familial mobility (4 children). In the end, five children were included in the sample reported here. There were three girls and two boys (mean age at first test: 49.4 months, sd : 6.2 months).

For information on individual participants, see table 4.

**III.II.1.2. Materials**

I used classic belief attribution transfer tasks (like Sally/Anne), each involving one or two characters, a hidden object, and two or three hiding locations. A small stage (50cm x 25cm x 25cm) was sometimes used, and other times the stories were demonstrated directly on a classroom table. Materials changed for each task and for each session so that children were not able to build a stimulus-response association for the tasks. See appendix 1 for example variations on the stories. Each session was videotaped for the record.

### III.II.1.3. Procedure

I visited the preschool in the morning or the afternoon, and asked children to play a few games on an individual basis with the same female experimenter every time. Every testing session took place in the same quiet room of the preschool and lasted about 15 minutes. Testing sessions occurred approximately once per month in accordance with the schools' schedule (mean interval between sessions: 39.4 days, sd: 11.8 days).

On the first visit, I only tested children's vocabulary level with the Peabody Picture Vocabulary Test- Revised (Dunn & Dunn, 1981). I did this because I wanted to ensure that children who participated were functioning at least at their age level and any difficulty with our tests could not be attributed to general language delays.

After the first session when I administered the PPVT-R, children received the same three tasks on every visit, with superficial variations in the story line and the props used (e.g., Bert leaves school while Ernie moves his books to a new place; Anne goes to get food for her dog who runs and hides under the bed). The tasks that children received on each visit were: false belief- unexpected transfer, false belief- unexpected transfer "look first", and true belief- unexpected transfer. In each story, a character, say Sally, is playing with an object and puts it away in one hiding location, before leaving the scene. Then, either a) Sally returns to see the object being transferred to another hiding location (true belief scenario), or b) Sally returns after the object has been transferred to another hiding location (classic false belief and false belief "look first" scenarios). See appendix 1 for examples.

The tasks' order was counterbalanced across visits. Each task included two control (reality and memory) questions, and two test questions. All four questions were asked at the end of the story, starting with one test question, followed by two control questions, and finally the second test question. The order of test questions was counterbalanced across sessions, as was the order of the control questions. Counterbalancing the order of the questions, and asking control questions would allow us to see whether children were perseverating in their answers.

At the end of each session, children received three stickers to reward them for their participation.

| Participant | Age at first session (in months) | Number of sessions | Mean interval between sessions (in days) (standard deviation) | Starting vocabulary age equivalent (in months) (percentile) |
|---|---|---|---|---|
| S1 | 49 | 18 | 37.5 (11.3) | 48 (47th) |
| S2 | 42 | 16 | 44.2 (15.5) | 26 (2nd) |
| S3 | 51 | 5 | 35.5 (9.7) | 49 (45th) |
| S4 | 46 | 5 | 43.3 (7.4) | 46 (50th) |
| S5 | 46 | 11 | 35.3 (5.8) | 45 (53rd) |

Table 4. Demographic and session information for children in cohort 1.

**III.II.2. Results**

**III.II.2.1. Coding**

I coded the responses for each child for each test question as failing (0) or passing (1), chronologically. For the false belief test questions, the child received 1 if they answered with that the character will look in the original location for the object, indicating the character will be mistaken about where the object actually is. The child received 0 if they indicated the character will look in the current location of the object, indicating the character will know where it is even though they did not witness the transfer. For true belief scenarios, the child received 1 if they predicted the character will look in the object's current location, and 0 if they predicted the character will look in its original location. There were three responses per session for the classic false belief questions: two for "think" and one for "look". I grouped these questions together under the "classic false belief" heading because these questions are equivalent in terms of cross-sectional group average rates of passing according to Wellman, Cross & Watson's (2001) meta-analysis. There was no reason to believe there should be a difference in the questions longitudinally. There were two responses per session for the classic true belief questions: one "think" and one "look". There was one response per session for the "look first" action prediction question based on false belief attribution. This gave us a string of 0's and 1's in the order in which the responses occurred.

**III.II.2.2. Cumulative records of performance for each individual on each task**

For each single case, I looked at the cumulative chronological records of responses on a task-by-task basis. The cumulative record was constructed by inputting the string of 0's and 1's for each individual's performance on each task seperately into Matlab. Each time the child passed a trial, the cumulative record increased by 1. Each time they failed, the cumulative record increased by 0. Thus, if the child performed at a consistently, the cumulative record would have a constant slope. Any change in the rate of passing would imply a change of slope in the cumulative record.

**III.II.2.3. Gallistel's change point algorithm for individual records**

I plotted the cumulative record of responses for each task for each child individually, to obtain acquisition curves. Then, using the Gallistel algorithm implemented in Matlab, I asked whether there were significant change points in the acquisition curves. This feature of each individual's developmental curves would allow us to address our questions about the profile (rate and shape) of acquisition within and between tasks for a given individual. The algorithm was borrowed from researchers in animal learning (Gallistel, Fairhurst & Balsam, 2004; Papachristos & Gallistel, 2006). If performance changes abruptly, then there is no need for such an algorithm to tell us whether there was a change. I would see it plainly in the data. Still, the algorithm can tell us whether there is a statistically significant change in performance. However, as I explained in the review of literature in Chapter I, I expected there to be periods where a child was neither at floor nor at ceiling. When such is the case, it would be necessary to use statistics to determine whether there is a significant change in performance over many data points.

The Gallistel algorithm operates in an iterative fashion on each point in a curve, starting from the origin. At each point, it asks whether there is a significant difference in the two cumulative distributions on either side of it. In other words, it asks whether the slope of the cumulative record changes significantly at this point on the curve. If the answer is no, the distributions are not statistically different in the binomial probability (of 0's and 1's). If the answer is yes, then it denotes that point as a change point at a given alpha level. (4) The algorithm then iterates to the next point in the cumulative record.

I ran the chronological string of 0's and 1's for each task for each individual separately through the algorithm. It is possible to select the sensitivity criterion when running the algorithm, so that change points can be detected at various alpha levels. The results reported here are all for criteria of $p \leq .05$. The product of the algorithm is a figure with two graphs, representing the data in two ways. In both graphs, trials are shown horizontally along the x axis, while performance is quantified vertically along the y axis. The graph on the top always represents performance in terms of the cumulative record of responses. This is the record over which the algorithm operates. The graph on the bottom shows the mean which best represents the cumulative record. Whenever the algorithm detects a change point in the cumulative record's slope in the graph on top, this is reflected in a change in the mean performance in the figure on the bottom. Whenever the mean in the graph on the bottom is the same throughout the entire period of testing, it means that there was no significant change in the cumulative curve representing performance over time.

Before we proceed, a note on the figures. They are the crux of the analyses here. It was not possible to standardize the axes in the cumulative records of responses (top panel of each figure). The x-axis depended upon how many trials an individual received. It varied from about five to about 60. The y-axis depended upon the total number of correct answers given throughout testing. It varied from one to about 60. Because y-axes in the cumulative record cannot be standardized, one is invited to consult the bottom panel of the figure where the mean response rate is presented. Again, x-axes for mean response rate depend on the number of trials. However, the y-axes could be standardized because they represent a rate ranging from zero to one (consistent failure to consistent passing). For ease of reading, the y-axis of the mean response rate (bottom panel) is always presented from 0 to +1.2. This way if the mean rate of correct responses was one the line would not be confused with the border of the figure.

I first report the results on an individual basis, and then I look across individuals at the group profile.

### III.II.2.4. Individual results

For each individual I noted the rate of passing each task overall, and before and after the change point if there was one. Then, I compared these characteristics of the curves between tasks for each individual child to dress a sequence of development for the three tasks within individual children.

**III.II.2.4.1. Participant 1**

See figure 5.

This child was tested every 37 days on average from 49 to 71 months of age.

True belief

Performance on the true belief- expected transfer task was high from the beginning. Overall mean performance was 86%. However, two change points were detected: one downward shift in the rate of correct responding from 100% to 16% at trial 20 (corresponding to session 10), and one upward shift back to near ceiling (from 16% to 100%) three sessions later.

False belief- unexpected transfer “look first”

The entire period of testing was characterized by the same rate of 33% passing the “look first” action prediction question. Although the algorithm detected no significant change in performance with alpha = .05, it does seem that there is a change at trial 10 (session 10), when the child first passes and then continues to pass afterwards at a rate of 62.5%. The last four trials (trials 15-18) are successful.

False belief- unexpected transfer

Overall, performance on the classic false belief task was 30% correct. The algorithm detects a change in performance, from 8% to 76% correct, at trial 38 (session 12).

Between tasks

This child’s performance on the true belief task was better earlier than on the false belief “look first” task, which was in turn better earlier than on the classic false belief task. Change points in the true belief and the "look first" task occurred at the same time (in opposite directions). Change points in true belief performance (and in "look first" performance, though not statistically significant) preceded the change point in classic false belief performance by two sessions.

**III.II.2.4.2. Participant 2**

See figure 6.

The child was tested every 44 days on average, from 42 months of age to 68 months of age.

Within tasks

True belief- expected transfer

Although the three-and-a-half year old had a vocabulary age equivalent of 26 months at the beginning of the study, in the $2^{nd}$ percentile for the age group, the child's performance on true belief attribution questions was very good (100% correct).

False belief- unexpected transfer "look first"

The child's performance did not change during the course of the study, with 13% correct answers to the "look first" action prediction questions overall.

False belief- unexpected transfer

By the end of the study, the child had only given 4% correct answers.

Between tasks

True belief "think" and "look" questions were mastered by this child from the beginning of testing, despite low language abilities (100% correct throughout the study in true belief scenarios). Performance on false belief tasks was in stark contrast. The false belief- unexpected transfer "look first" question elicited a higher percentage of correct responses overall from the child than the classic false belief task. The child's performance in both false belief tasks is not confused or random. There is a bias to select the "reality" answer in false belief attribution questions.

**III.II.2.4.3. Participant 3**

See figure 7.

This child participated in six sessions every 35 days, from 51 to 58 months.

Within tasks

True belief- expected transfer

Performance in the true belief condition throughout our study was represented by an overall mean of 50% passing. The algorithm detected no significant change in the cumulative record of responses. It appears, though, that performance was high at first (75% correct over the first eight trials), but later declined. The last four trials (sessions five and six) were unsuccessful.

False belief- unexpected transfer "look first"

Performance on the false belief- "look first" task was near ceiling from the beginning of testing, characterized by an overall mean of 83%.

False belief- unexpected transfer

At trial 10 (session 4), the rate of passing increased significantly from 10% to 62.5%. Overall, the child gave 33% correct responses on the unexpected transfer false belief attribution task.

Between tasks

Performance on the true belief task and the "look first" task were comparable at first. Later, performance on the true belief task declined. Performance on the false belief- unexpected transfer task appears to lag behind performance on the false belief unexpected transfer "look first" task. If I only looked at the overall mean throughout the period of testing for each task, regardless of the chronology of the responses, it would appear that the "look first" task was easiest (83% correct), followed by the true belief task (50%) followed by the classic false belief questions (33%). This is a case where using our change point algorithm on a cumulative record of performance over time more accurately reflects development than an overall mean would.

**III.II.2.4.4. Participant 4**

See figure 8.

This child was tested every 43 days on average, from 59 months to 67 months of age.

Within tasks

True belief- expected transfer

Performance in the true belief task was high (92% correct responses overall). There was no significant change in performance.

False belief- unexpected transfer "look first"

Performance on the "look first" version of the false belief task was consistently high for this child, and there was therefore no point of change in the performance. Overall, the child gave 83% correct answers to the "look first" question.

False belief- unexpected transfer

Performance changed at trial 9 (session 3), from 11% correct to 62.5% correct. Overall, this child's performance on the classic false belief task is characterized by a mean of 33% correct.

Between tasks

The same profile of acquisition seems to emerge here too. The child passed true belief attribution questions at the highest rate overall. The "look first" task is passed at a high rate as well (though not as high as for true belief scenarios), while performance on the classic false belief "think" and "look" questions are only passed one in three times.

**III.II.2.4.5. Participant 5**

See figure 9.

This child is tested every 35 days on average, from 46 months to 58 months.

Within tasks

True belief- expected transfer

True belief scenario "think" and "look" questions are passed 95% of trials, throughout the period of testing. There is no significant change in the child's performance.

False belief- unexpected transfer "look first"

The child's performance showed a consistent bias throughout the period of testing to report the wrong answer to false belief questions in the look first scenario. Overall, only 9% of responses were correct.

False belief- unexpected transfer

During the period of observation, this child only passed 3% of the classic false belief "think" and "look" questions.

Between tasks

This child's performance on the true belief task was at ceiling, while performance on both the classic false belief-unexpected transfer task and the false belief- unexpected transfer "look first" task were at floor throughout. Strictly speaking, false belief "look first" was passed at a higher rate throughout the study than classic false belief tasks.

**III. II.3. Discussion**

Our goal was to use a new method to investigate children's development in theory of mind. I looked at individual cumulative records of performance and I scanned the curves for significant change points in the slope. Having this information about the slopes of the curves for each task and each child gave us a reliable, quantitative approach to describing developmental profiles. Now, I look across single cases at the profiles I drew, and I ask whether there were any commonalities or differences between the profiles of development for individual children.

### III.II.3.1. Individual variability

I was able to replicate and extend findings from other similar studies of individual children's development (Rein, 2005; Flynn, O'Malley & Wood, 2004; Flynn 2006). Individual results showed variable patterns of developing theory of mind, as measured by false belief attribution. What I couldn't have described from the group averages were the various individual patterns I observed. Saying that most children pass at five and fail at three could have meant many different things about individual development. I found that in fact there was no universal pattern either within an individual across tasks or across individuals within tasks. Children show a variety of patterns of individual development on each theory of mind task they received. Children in our individual analyses were in their fourth year when they began to pass consistently. Group averages suggested that young three-year-olds are biased to consistently fail the false belief task, while older three-year-olds' performance is closer to chance. Group averages also suggested that children who are older than four perform consistently above chance. This could mean that all children begin to pass the false belief attribution task consistently around the age of four, and that half of the children who are nearing their fourth birthday will pass consistently, while the other half will fail consistently. However, in light of the individual cumulative records of performance given in our study, together with the individual records observed by Flynn (2006) and Rein (2005), it seems that the group averages represent overlaps in single case records of increasing rates of passing. The rate of passing for some children increases quickly, while for other children it increases slowly.

**III.II.3.2. Controls**

I further explored these individual patterns in order to see whether they would be the same in other belief attribution tasks. I wanted to be assured that any variability I observed in performance on our tasks was not due to processing limits unrelated to their theory of mind. I added a true belief task, and showed that variability in false belief performance over time was not found in true belief scenarios.

High rates of passing our control questions showed that children were not simply too young to follow the task. Also, age-appropriate vocabulary levels illustrated that the children in our study (except S2) had a good grasp of the language in the task. In addition, I included a true belief- expected transfer task, which is analogous to the false belief attribution task.

I expected that performance on true belief tasks would be less variable than performance on false belief tasks. Indeed, I expected that even the youngest children in our study would show ceiling performance in the true belief task. This is what I found, with one exception. I observed a regression in performance by S1 on the true belief task which was mastered early at a high level. This regression occurred in temporal proximity to change points on the false belief tasks (within 10 weeks when period preceding change was 53 weeks).

It encourages us to take the within- and between-individual variability observed on the false belief tasks seriously. The profiles of development I drew for false belief tasks reflect changes in processing which I did not see at the same time (in the same individual) in true belief tasks.

Traditionally, I could look at the mean performance over a group of responses to see which one was higher. However, I have additional temporal information here that allows us to give a more detailed picture of development. The overall rate of passing is misleading in cases where a significant change in performance occurred, because it represents a mean over time. A child may have an overall rate of 30% passing in task A, and a rate of 60% passing in task B, but this could mean different things about the underlying curve. For example, the overall rate of 30% could mean the child passed one out of three trials for several months, or never passed, then consistently passed for the last third of the trials. Which case we find ourselves in will affect how we judge the relative difficulty of tasks, and the developmental relation between the abilities they measure.

### III.II.3.3. Sequence of development

I found that two subjects (S2 and S5) only gave one or two correct responses to each false belief task. Nonetheless, for both of these children, false belief "look first" was passed at a higher rate than the classic false belief task. There were two subjects (S3 and S4) who passed the false belief- "look first" task at a high rate from the beginning of testing, while their performance in the classic version started out poorly, but improved significantly after a few sessions. These children had difficulty with the classic version for longer. This means that their performance on the standard version lagged at least four months behind their performance on the "look first" version. Finally, S1, like S3 and S4, showed a high rate of correct answers in the "look first" version before a high rate of responding in the classic version of our false belief attribution tasks.

In sum, all children passed the true belief task at overall high rates throughout the study. Children either failed both false belief tasks or they passed the false belief "look first" task before they passed the classic false belief task. This suggests that "look first" taps at least some processes that are distinct from the classic false belief task, and that it relies on abilities that develop earlier. This sequence of development revealed in our individual analyses is consistent with the sequence of development suggested by cross-sectional group averages (Siegal & Beattie, 1991; Surian & Leslie, 1999).

**III.III. Cohort 2: Exploring a role for executive functions in theory of mind development**

I showed with our longitudinal data from cohort 1 that paths of change are variable across individuals. Some children changed quickly, while others changed slowly. This variability was not an artifact of repeated testing, as our control true belief condition did not produce the same type of variability. Having a picture of what change looks like lays the ground work for us to ask what motivates these changes in performance. Leslie and colleagues' hypothesis is that theory of mind development depends upon sufficient executive resources being available to the child (Surian & Leslie, 1999; Leslie & Polizzi, 1998; Friedman & Leslie, 2004a, 2004b; Friedman & Leslie, 2005). Surian and Leslie (1999) suggested a mechanism to explain the facilitation effect of adding "first" to the look question by a salience account of what is required of children in the false belief task. That is, the authors call on the ToMM-SP account of belief attribution, whereby a specialized modular system determines what belief contents are available (Theory of Mind Mechanism) and another system selects the answer among the candidates (Selection Processor). According to this model, true belief contents act as a default, so that when reasoning about a false belief scenario, children must overcome the strong, but wrong, true belief response. If they have sufficient capacity to inhibit the wrong answer and select the correct answer, they will be successful in the false belief task. If they are not able to muster the selection strength required to overcome the true belief candidate, then this is the belief content that will be attributed to the character in the story. Leslie and colleagues have found data consistent with the ToMM-SP model of reasoning about beliefs (Surian & Leslie, 1999; Leslie & Polizzi, 1998; Friedman & Leslie, 2004a, 2004b; Friedman & Leslie, 2005). Other researchers too have suggested that the link between theory of mind development and executive processing resources can be useful for understanding the deficits in autism (Russell, Sharpe, Mauthner & Tidswell, 1991; Ozonoff, Pennington, & Rogers, 1991). Indeed, in our first chapter I reviewed some of the correlational evidence for a relationship between theory of mind and executive processing in normal preschoolers.

This is, however, still a controversial claim. Some have questioned whether executive functions are a unitary construct that is measured by both theory of mind tasks and standard executive function tasks (Leslie, German & Polizzi, 2005). Others have questioned whether the relation between theory of mind and executive functions is a direct causal one rather than mediated through general maturation. It could also be that the causal relation runs in the opposite direction. According to this view, theory of mind development, i.e., appreciating people as mental agents, understanding the connection between mental states and actions, etc., is a necessary component of executive functions. This view holds, then, that theory of mind should develop as a pre-requisite for executive functions, rather than the other way around. Flavell (1969) calls the type of developmental relationship under investigation regarding theory of mind and executive functions a “non-implicative mediation”, meaning that an empirical investigation, rather than a purely logical one, is the only way to resolve the question of the sequencing between these abilities.

A few studies have addressed the question of the developmental relation between theory of mind and executive functions in a longitudinal manner. Hughes (1998) used cross-lag correlations to reveal that theory of mind at time 2 is predicted by executive processing at time 1, but the reverse did not hold. Carlson, Mandell, and Williams (2004) tested children twice, at 24 months and at 39 months. These children’s executive functions scores at time 1 predicted theory of mind scores at time 2, even when executive function scores at time 2 were factored out. In contrast, the theory-of-mind scores at time 1 were not correlated with executive function scores at time 2.

Flynn, O’Malley and Wood (2004) gave children theory of mind questions and executive processing tests, and followed their development through six sessions over six months. Under a moderately strict criterion of passing (at least 2/3 correct for false belief and 12.5/ correct for the hand game and a maximum of one commission error on the lights game) 16 children passed the inhibition hand game while failing the theory of mind tasks, while three children showed the opposite pattern. Twenty children passed the inhibition lights game while failing theory of mind, and no children showed the reverse pattern. Overall, these results are consistent with the suggestion that children will not pass theory of mind tasks until they have gathered sufficient executive processing strength.

Our aim was to explore the developmental relation between executive functions and false belief attribution in individual preschoolers. To measure these developing abilities, I used the classic false belief-unexpected transfer task, and I added the classic false belief- unexpected contents task. In his microgenetic study of individual theory of mind development, Rein (2005) found that seven children out of 11 whose performance changed significantly during the period of testing showed higher rates of passing the false belief- unexpected transfer earlier than passing the false belief-unexpected contents task. Only four out of 11 children showed the reverse pattern. Flynn (2006) also found this sequence of development in her microgenetic study, while Wellman, Cross and Watson (2001) reported no significant difference in difficulty between these two theory of mind tasks in cross-sectional group averages.

I also used one of the four conditions from the game developed in chapter one as a measure of executive processing. I chose to administer the condition where children first received the incongruent rule, then the congruent rule, because this would test two different types of inhibitory capacity. Using the incongruent rule in the first block would require children to inhibit the prepotent responses of pointing to what they hear. In other words, they would have to select between two competing response options immediately available to them. In the second block using the congruent rule, children would have to inhibit a prepotent rule from a previous trial (or block). This time, the focus of the competition is between two separate rules, rather than two response options.

I expected that children would successfully use the incongruent rule before they would successfully complete the false belief tasks. This would support the case that executive processing, and in particular selection of the correct response in the face of an initially stronger competitor, is a large component of the false belief task. Consistent with experiment 2 of Chapter II, successfully applying the incongruent rule in the first block should develop some time between the ages of three and five. I did not expect the congruent rule to be related to false belief performance in a developmentally relevant way. Experiment 2 of Chapter II showed that three-year-olds and five-year-olds found it equally difficult to switch to the congruent rule in the second block.

**III.III.1. Method**

**III.III.1.1. Participants**

After obtaining permission from preschool directors, I sent recruitment materials to all children between 3;0 and 5;11 in each school. This age range was chosen to maximize our chances of observing change during the period of testing. Cross-sectional and longitudinal studies of theory of mind and executive functions in preschoolers indicate important changes occur in the preschool years. Eight children agreed to participate. There was one child who completed less than five testing sessions then changed schools, and one child who missed more than two consecutive sessions. Thus, only six children were included in the final sample reported here. There were three girls and three boys (mean age at first test: 51.7 months, sd: 6.9 months).

For information on individual participants, see table 5.

**III.III.1.2. Materials**

A small stage (50cm x 25cm x 25cm) was used for false belief- unexpected transfer stories. A familiar container (e.g., a cookie box) and contents (e.g., a sock) were used for false belief- unexpected container tasks. Materials changed for each task and for each session so that children were not able to build a stimulus-response association for the tasks. To assess inhibitory control, I used a new test of abstract executive functions in preschoolers, called the opposites task. This task required an 8x11 inch binder with laminated pages. Each page of the binder showed two pictures side by side, separated by a line. The pictures were approximately 4 inches in diameter. Some presented semantically opposed pairs of pictures (e.g., open and closed doors), while others represented variations along a relevant dimension (e.g., swimming and flying ducks). Each session was videotaped for the record.

**III.III.1.3. Procedure**

I visited the preschool in the morning, and asked children to play a game on an individual basis with the same female experimenter every time. Every testing session took place in the library of the preschool and lasted about 15 minutes. Testing sessions occurred approximately once per month in accordance with the schools' schedule (mean interval between sessions: 39.4 days, sd: 11.8 days).

On the first visit, I tested children's vocabulary level with the Peabody Picture Vocabulary Test- Revised (Dunn & Dunn, 1981). I did this because I wanted to ensure that children who participated were functioning at least at the level of three-year-olds.

After the first session when I administered the PPVT-R, children received the same three tasks on every subsequent visit. The tasks that children received on each visit were: false belief- unexpected transfer, false belief- unexpected contents, and the opposites task.

Theory of mind tasks- unexpected transfer and unexpected contents

To assess theory of mind development, I used classic unexpected transfer and unexpected contents false belief attribution tasks (see appendix for details). Each false belief task included control questions, which preceded the test question. Children could receive one point for a correct answer to the false belief – unexpected transfer task, and one point for a correct answer to the false belief- unexpected contents task. Thus, their aggregate theory of mind score could vary between 0 and 2.

Inhibitory control task- incongruent rule and congruent rule

In the opposites game, children first were trained with the incongruent “point to the other picture” rule, then played the game for six trials with six different pairs of pictures They could receive a score between 0 and 6 for this first block of trials. This first block of trials would test children’s ability to inhibit a prepotent response, in order to select the correct picture. At the end of those six trials, the rule was switched and children were trained with the congruent “point to what I say” rule, then played the game for six trials with six new pairs of pictures. In this second block, the currently relevant congruent rule should be easy, except that children had to ignore a previously relevant (incongruent) rule. This, children had to inhibit a prepotent rule in the second block, in order to select the correct response. They could receive a score between 0 and 6 for this second block of trials.

I varied some aspects of the tasks between sessions, as stimulus-response pairing could be an issue due to repeated testing. I operated superficial variations in the story line and the props for the two false belief tasks. The order of stimuli in the opposites task (i.e., the order of the picture and word pairs the child was exposed to) was reversed for half the sessions. The tasks’ order was also counterbalanced across sessions. See appendix 2 for examples.

At the end of each session, children received three stickers to reward them for their participation.

| Participant | Age at first session (months) | Number of sessions | Mean interval between sessions (days) (standard deviation) | Starting vocabulary score (age equivalent) (percentile) |
|---|---|---|---|---|
| S6 | 42 | 11 | 47.8 (32.9) | 62 (91st) |
| S7 | 62 | 14 | 41.7 (15.9) | 57 (37th) |
| S8 | 53 | 14 | 32.1 (11.4) | 61 (77th) |

| | | | | |
|---|---|---|---|---|
| S9 | 51 | 6 | 39.6 (23.3) | 55 (68th) |
| S10 | 55 | 6 | 32.6 (11.4) | 46 (19th) |
| S11 | 47 | 11 | 47.8 (19.8) | 60 (88th) |

Table 5. Demographic and session information for cohort 2.

### III.III.2 Results

I coded the responses for each child for each test question as failing (0) or passing (1), chronologically. This gave us a string of 0's and 1's in the order in which the responses occurred. As for cohort 1, I plotted the cumulative record of responses for each task for each child individually. For each cumulative record I measured the change point(s). Individuals' results are reported before looking across individuals at the group profile.

Individual results

For each individual's performance, I looked at the cumulative chronological record of responses on a task-by-task basis. There were six responses per session for the "incongruent" rule, and six responses per session for the "congruent" rule. There was one response per session for the false belief – unexpected transfer task (classic "think" question). There was one response per session for the false belief – unexpected contents task (belief attribution to "other"). The cumulative record for each task for each individual was analyzed using Gallistel's change point algorithm (Gallistel, Fairhurst, & Balsam, 2004). All the results reported here are for criteria of $p \leq .05$.

#### III.III.2.1. Participant 6

See figure 10.

This child was tested on average every 48 days from 42 months of age to 59 months of age.

Incongruent Rule

Overall, there were 59% correct answers using the incongruent rule. There was one significant change point in performance, at trial 29 (session 5). This marked a progression from 10.3% passing to 97% passing.

Congruent Rule

When considering the congruent rule, this child's performance was stable at 89% correct.

False belief- unexpected transfer

Performance changed at trial 5 (session 5) from 0% to 83% correct. The overall mean for false belief attribution in the unexpected transfer task was 45%.

False belief- unexpected contents

This child's performance in the false belief unexpected contents task did not change throughout the period of testing, remaining at an overall level of 28% correct.

**III.III.2.2. Participant 7**

See figure 11.

This child was tested every 42 days on average from 62 months to 81 months.

Incongruent Rule

There was no significant change in performance throughout the 19 months of testing. Performance reached 94% correct overall.

Congruent Rule

Performance on the congruent rule was initially low (30%), but changed significantly at trial 10 (session 2) to a mean of 95%. If I looked only at the overall mean, the child succeeded on 87% of the trials.

False belief- unexpected transfer

Performance in the unexpected transfer task was at ceiling (100%).

False belief- unexpected contents

Performance in the false belief unexpected contents task did not change throughout the study, remaining at 79% correct.

### III.III.2.3. Participant 8

See figure 12.

I tested this child from 53 to 67 months of age, every 32 days on average.

Incongruent Rule

The algorithm confirmed what the eye could see: there was a significant change in performance on trial 11 (session 2), from 0% to 96% passing. Overall performance throughout the 14 months of testing was represented by a mean of 80%.

Congruent Rule

This child missed only one trial out of 84, yielding a mean score of 99% correct.

False belief- unexpected transfer

Performance changed significantly on trial 5 (session 5) from 0% passing to 100% passing.

False belief- unexpected contents

Only 14% of the trials for this task were successful. There was no significant change in performance throughout testing.

### III.III.2.4. Participant 9

See figure 13.

This child was tested from 51 months to 59 months, every 40 days on average.

Incongruent Rule

This child's performance on the incongruent rule maintained a high level of success (92% passing throughout).

Congruent Rule

Only one wrong answer was given out of 36 trials (overall mean correct: 97%).

False belief- unexpected transfer

There was no change in performance, with a perfect score throughout the period of testing (100% correct).

False belief- unexpected contents

There was no change in performance throughout the 6 sessions. The child succeeded on every single trial (100% correct).

**III.III.2.5 Participant 10**

See figure 14.

I tested this child starting at 55 months every 33 days, until the age of 61 months.

Incongruent Rule

Performance on the incongruent rule was near ceiling during the entire study. Of 36 trials on the incongruent rule, this child failed only 3 (92% passing).

Congruent Rule

This child's performance using the congruent rule was near ceiling (97% correct).

False belief- unexpected transfer

Every trial in the unexpected transfer false belief task was successful for this child. They obtained an overall score of 100% correct.

False belief- unexpected contents

Performance on the other false belief task, requiring attribution of a false belief about the unexpected contents of a box, was at chance. Throughout the testing period, the child only managed 50% correct answers.

**III.III.2.6. Participant 11**

See figure 15.

This child was enrolled in the study from 47 months to 64 months of age, and was tested every 48 days on average.

Incongruent Rule

Performance on the incongruent rule did not change, with an overall mean of 93% correct.

Congruent Rule

The algorithm indicated two significant changes in performance on the congruent rule. There were therefore three distinct periods in the child's performance. The first change point, at trial 25 (session 5) marked a regression from 92% to 0%. Then at trial 29 (session 5) performance returned from floor to ceiling (literally 0% to 100%).

False belief- unexpected transfer

This child's performance did not change, remaining near ceiling throughout (100% correct).

False belief- unexpected contents

Performance on the unexpected contents task was also very good (91%) throughout the 11 sessions.

**III.III.3. Discussion**

Data from the second cohort of children in our longitudinal studies helped us extend our investigations of children's developing theory of mind. Due to random sampling, the children in this second cohort were older than in cohort 1, and their performance on the classic false belief unexpected transfer task was overall very high (mean age at start cohort 1 = 46.8 months; mean age at start cohort 2 = 51.7 months). I included a second false belief task, the unexpected contents task, and I found that children's performance on the unexpected contents task lagged behind performance on the unexpected transfer task. This confirmed the relative difficulty of the tasks described by Flynn (2006, p. 645), and replicated by Rein (2005). This microgenetic result, obtained in three separate cohorts using our method, leads us to wonder why cross-sectional group averages did not reflect this relative task difficulty between the unexpected transfer and the unexpected contents tasks in Wellman et al.'s (2001) meta-analysis.

The main purpose of the second cohort was to utilize this new method to establish developmental relations between tasks, within individual children. I administered a test of children's inhibitory capacity alongside our false belief tasks at every session. The picture that emerged was one of a pair of inhibitory abilities: the incongruent block of our inhibition task tested children's ability to overcome a tendency to point to the word they heard. The congruent block required children to switch rules (away from the incongruent rule) and simply point to what they heard. In terms of the developmental relation between theory of mind and inhibitory processing, I saw that changes in the incongruent block were yoked with changes in theory of mind, but changes in the congruent block were not. Children whose performance in the first incongruent block improved significantly also showed a significant improvement in theory of mind performance either in the same session or in a subsequent session. When a child's performance in the second congruent block changed significantly, there was no such echo in theory of mind performance. It seems, then, that our two blocks in the inhibition game tested different types of inhibition. Within trial inhibition was apparently related to theory of mind performance, while inhibition of a previously relevant rule was not apparently related to theory of mind performance.

**III.IV. General discussion**

I successfully implemented a new longitudinal method of studying development using intensive testing and a change point algorithm to detect changes in the rate of correct responses. Our results encourage us to continue to use this new method, despite its heavy logistical requirements. An optimal use of this new method would include careful consideration of three issues: the window of observation, the rate of sampling and the number of participants at the initial recruitment.

a) Window of observation

Ideally children would be recruited in an optimal age range and followed long enough to observe individual patterns of development on all of the measures one chooses to take. In practice, this is very challenging. The best efforts can be made towards recruiting a large number of individuals within the range of ages where change is expected, so that attrition is not a problem for final analyses.

b) Rate of sampling

In addition, I found that in some cases a change in performance was apparent to the naked eye whereas the algorithm did not detect a change point. In these cases raising the alpha level (relaxing our criterion) did not help to bring out a change point. What did help to bring it out was adding another testing session. The algorithm found a change point in the record once there was a larger data set.

c) Number of participants

In addition to analyzing individuals' patterns of development within and across tasks, this new method could be used to compare patterns across individuals. Just as I subjected the individual records to inferential statistics using the Gallistel algorithm, with enough participants we could use inferential statistics across children. In future studies I will aim for a final sample size of at least 20 children so that conclusions across individuals can be drawn on an inferential statistical basis. If there were enough children in our final sample, we could use inferential statistics to conclude whether one pattern is significantly more common than another across individuals. This type of analysis could be done not only for the patterns of development on a given task, but also for the sequences of development between tasks.

I was interested in two fundamental developmental questions. I wanted to see what development looks like on a single case basis, for each task. I could look at change points within a child's performance on each task, and take note of the rates of passing before and after any change point. This gave us information that overall means alone would not have given us. I also wanted to compare development across tasks for each individual I worked with, to see whether I could establish a sequence of development between the tasks. Determining a sequence of development between tasks within individuals which holds across individuals would give us hope of inferring a causal relationship between these abilities, better than a correlational method would.

Concerning the shape of single case developmental profiles, I replicated Rein (2005), Flynn, O'Malley and Wood (2004), and Flynn's (2006) findings that there are various individual paths to mature performance in theory of mind tasks. I found instability in individuals' developmental profiles. I also found variability across individual profiles. In addition to asking what the path and rate of change was for each individual's cumulative record on each task, I compared each child's performance on one task to their performance on other tasks. Data are consistent with the suggestion that reducing the executive requirements of a task allowed children to pass that task earlier. The sequence I observed between tasks within individuals in cohort 1 was consistent with this: children's performance on the false belief "look first" question was better earlier than their performance on the classic false belief "look" and "think" questions. Further, in cohort 2, I showed that executive functioning could be a pre-requisite for theory of mind. Children's performance in our incongruent block of the executive function game changed before performance on theory of mind changed.

I never had more than two change points, while Papachristos and Gallistel (2006) found for some mice that there were many change points. Many factors could explain why we might find more or less change points in a given cohort. It could be that the changes in performance that a child's cognitive structures undergo in our tasks are simply less numerous than the changes a mouse undergoes in a foraging task. Whether developmental change is quantitatively similar across domains is an open question. It is also possible that there were in fact many small changes in the rate of passing our tasks, but these changes are not detected by our manner of observation. It is always possible that I chose to look into a time window that was too early or too late regarding a given individual's development, or that I simply didn't have enough data for our statistical test to detect change.

Initially the age group for recruitment into our longitudinal paradigm was chosen based upon age differences in the cross sectional literature. This age group was perhaps even a bit on the older side. In cohort 1, children were at ceiling on the true belief task from the beginning (though not on the false belief tasks). In cohort 2, children were at ceiling on the false belief unexpected transfer task from the beginning (though not on the false belief unexpected contents task). Ideally, we would observe children throughout the course of change on every task we administered. This may be impractical because apparently change is very slow. If I wanted to observe change in true belief and in false belief in the same child, I might need to test them for two or three years at least once per month.

Future studies will tell us whether the shape of the increase in an individuals' rate of passing (whether abrupt or gradual) is a function of the frequency of sampling, feedback, or other factors. In particular, it would be interesting to ask whether feedback changes the shape of the developmental curves, or if feedback accelerates acquisition, or both. Amsterlaw and Wellman (2006) found that feedback (witnessing where a character actually searches when they have a false belief) facilitated improvements in a group of preschooler's performance in various theory of mind measures, compared to a control group. Also, the mice in Papachristos and Gallistel's (2006) study received feedback in the form of food for successful trials. In both of these cases, it is possible that the rate of acquisition was affected by feedback.

## Chapter IV

### IV.I. Learning about development from cross-sectional group averages and longitudinal individual profiles

The questions we ask of development are often limited by the methods with which we attempt to answer them. In the past couple of decades, empirical investigations have charted the waters of children's developing understanding of people as mental agents to a great extent. We now know that when we ask young preschoolers to verbalize a point of view held by a character in a story, they tend to show an egocentric bias, focusing on their own perspective more easily than on someone else's. It is now a well accepted fact that preschoolers' ability to reason about mental states, such as a person's belief about where an object is located, changes within a couple of years from an egocentric bias to an appreciation of perspectives different from their own. In contrast to three-year-olds, five-year-olds tend to be capable of reporting that a character's belief is false, when the character did not witness a crucial turn of events. Group averages also tell us that the developmental timeline is relatively stable across cultures and socio-economic backgrounds. At the same time, group data hint at important inter-individual variability. Yet, current theories of developing theory of mind do not make specific predictions about what individual development should look like. Do all children follow the same developmental trajectory to attain mature theory of mind? Does each child undergo abrupt change in their performance on verbal theory of mind tasks, or is change gradual and non-monotonic? These basic developmental questions are not addressed by current theories of development. In fact, the most popular methods employed in the study of development do not allow us to ask those types of questions, because group-averages do not say anything about inter-individual variability and intra-individual instability. Two questions are left relatively open by group –average learning curves: a) what do individual profiles look like, which underly the group averages?, and b) what (if any) cognitive processes bear a causal relation theory of mind development? These questions about the cognitive architecture of the mind are both fundamental and difficult to explore empirically. In this dissertation I asked whether changes in preschoolers' mental state reasoning are motivated by changes in executive processing using both a cross-sectional group approach and a longitudinal single case approach.

## IV.II. Variability in theory of mind development in individuals and its relation to executive functions

### IV.II.1. A new measure for executive functions

In Chapter II I discussed the types of executive functions which might be related to theory of mind development. Adding to pre-existent measures of executive functions, I developed a new test for abstract inhibition in preschoolers. I tested preschoolers' ability to use an abstract rule in a game where they were required to inhibit their prepotent tendency to match a stimulus and a response. When our new test of abstract inhibition was given to groups of preschoolers, I revealed a developmental trend, from three-years-old to five-years-old, of increasing success at inhibiting the strong, but wrong, response. There was no such developmental trend for another type of inhibition I tested, that is, inhibiting a previously relevant rule, when the new rule was easy. However, when the new rule was difficult, younger preschoolers showed higher switch costs than older preschoolers. Finally, another group of children completed theory of mind tests and our new tests of inhibition. In this group, inhibition of a strong, but wrong, response was significantly correlated with theory of mind performance, even when chronological age and verbal age were factored out. This type of inhibition correlated more strongly with false belief-unexpected contents task performance than with false belief-unexpected transfer task performance.

**IV.II.2. A new method for evaluating variability within and between individuals**

Our new approach using the Gallistel algorithm on microgenetic data differs from Siegler's approach to microgenetic data analysis in at least two ways: the type of data we can examine, and the type of conclusions we can draw about individual records.

First, our data are bimodal (pass or fail) whereas Siegler typically classifies children's responses in terms of strategies. I was not asking which strategy a child uses to arrive at the correct answer, and how strategies gain or lose favor over time in a child's approach to a problem. My question was simpler than that and I did not have categorical data that Siegler's method accomodates. I wanted to look at the dynamics of the *rate of passing* only, characterized as a string of 0's and 1's in the order the responses occured. The Gallistel algorithm offered a simple way to analyze the slopes of our cumulative records created from bimodal data.

The analyses I ran with the Gallistel algorithm could indeed also be run on other forms of data beyond bimodal data. For example, we could gather explanations from children as they reason about theory of mind and classify the explanations along various dimensions (from less "mentalistically" mature to more mature). Once we have established categories, we could assign points for each category of explanation they use. Imagine for example that a child explains a character's behavior of looking for an object in a particular location by the fact that the object really is there. We might deem this a "reality-based" explanation, and assign 1 point because the didn't use mental states to make sense of the character's behavior. If the child has recourse to a "mentalistic" explanation, however, saying that the character looks in a particular location because they think an object is there, then we might assign 5 points. The particular categories we choose and the points we assign are taken here as examples, to illustrate that it would be possible to gather data from children in theory of mind that is not bimodal, and indeed people have done this. In these cases, we could use a Siegler-type analysis or a Gallistel-type analysis. However, I'm not sure how the Siegler approach handles bimodal data. The Gallistel algorithm handles bimodal data very well.

Another difference between the Gallistel analyses and the Siegler analysis is what they can tell us about the changes in performance we observe in an individual's record of performance. The Gallistel algorithm finds a change point in the cumulative record (a point where the slope of the record changed, reflecting a change in the rate of correct responses), using inferential statistics on an individual's performance. The Siegler method does not have a way of pinpointing change in an individual's record based on inferential statistics. Rather, descriptive statistics are used for individuals, and then individual data are turned back into group data for the purposes of inferential statistics. I believe the application of the Gallistel algorithm's inferential statistics to individual response records will allow us to look deep into individual's development. Finding a statistically significant change point is by definition very different from finding no statistically significant change point. Looking across individuals we have examined with our powerful statistics can tell us whether there is only one pattern that every individual follows, or whether there are children who are unusual. A Siegler-like method does not offer this sort of statistical power for individual records of performance.

In chapter III I employed the Gallistel algorithm to detect change points in individual development within and across theory of mind and executive function (inhibition) tasks. Cohort 1 completed false belief and true belief tasks. I found that long term variability in individual performance was present in false belief tasks, but not in true belief tasks completed by preschoolers. Thus, developing the ability to attribute false belief verbally in our paradigms was unstable and gradual. I also confirmed on a single case level what cross-sectional data suggested, that adding the word “first” to our action prediction question in the false belief scenario made the task easier for the children so that they passed at a younger age. Presumably, adding “first” to our question reduced the executive demands by making the right (false belief) answer a stronger contender in a battle against the default (true belief) wrong answer. The selection processor’s work load was lightened and children passed at a younger age.

Cohort 2 completed two false belief tasks and our new inhibition tasks. Performance on our inhibition measures was higher earlier than performance on our theory of mind false belief tasks. Also, I found that for most children the false belief- unexpected transfer task was easier (or at least was passed at a higher rate earlier) than the false belief- unexpected contents task. Cross-sectional group data from Chapter II Experiment 3 told us that our inhibition measure correlates more strongly with false belief- unexpected contents than with false belief- unexpected transfer. Together with evidence from Cohort 2, these findings suggest that the false belief- unexpected contents task is harder than the unexpected transfer task because of higher executive demands. Wellman, Cross and Watson (2001) did not find a difference in difficulty between these two tasks in their meta-analysis of cross-sectional group data. However, like us, in her microgenetic study Flynn (2006) found the unexpected transfer task was easier for children. Perhaps the difference in task difficulty between unexpected contents false belief attribution and unexpected transfer false belief attribution is small when children complete the tasks only once, but increases in a longitudinal context. One major difference between the paradigms is the degree to which the child's own knowledge (the true belief or "reality" content) is highlighted. In the false belief unexpected transfer task, the child witnesses a fictional story and answers questions about a fictional character's beliefs and actions. Longitudinally, the story unfolds with different fictional characters in a different context using different objects every time. In the unexpected contents task, the scenario is not fictional at all. The child sees an object unexpectedly removed from a familiar container, and must say what someone in their class would think is inside. Though the container and the contents, as well as the classmate to whom they must ascribe a belief, are different each time the child completes the task, the scenario itself is not fictional, and this could imply greater salience for the child's own (default) true belief. The selection processing difficulty introduced by highlighting the reality of the events (asking them to identify the familiar box; asking them to attribute a false belief to a classmate) in the unexpected contents paradigm may be very slight, compared to the unexpected transfer paradigm. However, when administered repeatedly over the course of many months, the difference in task difficulty may be compounded. When they are confronted with the unexpected contents scenario for the fourth or fifth time, children may know from the very start of the task that whatever the box appears to contain is not the contents. Thus, they begin the task by inhibiting (rightfully so!) the false belief content due to their own prior knowledge. This knowledge on the part of the child could interfere with the selection processor's ability to attribute the false belief ("apparent content") to another child.

This sequence of development between tasks within individuals gives a dynamic picture of changes while they are taking place, and informs the relation between these changes on an individual level. Furthermore, there were differences between children as well as instability within children. These data were predicted neither by theory-theory nor by modular theory. Factors which can explain the variability of development within tasks across individuals and within individuals across tasks should be explored.

## IV.III. The executive factor as explanatory mechanism for individual's developing theory of mind

### IV.III.1. Explaining variability within performance on a given task

Executive resources may be a factor which accounts for the individual profiles of change in theory of mind that I observed. One can imagine that as executive resources become increasingly available to the child, the probability of the child giving the correct answer will increase. According to the simplest form of this model, the shape of the individual curve in theory of mind is a function of the increase in executive resources. If executive resources increase slowly, the curve of performance in theory of mind increases slowly. If executive resources accumulate until they abruptly cross a threshold, then an abrupt change in theory of mind would be observed. These suggestions for explaining intra-individual instability on a task, as well as inter-individual variability on a task, are avenues for future research.

### IV.III.2. Explaining the sequence of development between tasks

Could the executive factor explain why some sequences of development between tasks are shared between individuals? In cohort 1 I found that tasks with lower executive demands elicited higher rates of passing earlier than tasks with higher executive demands. True belief attribution developed before false belief attribution in the classic unexpected transfer paradigm. Attributing a false belief in the "look first" paradigm developed before attributing a false belief in the classic paradigm. Also, when inhibitory capacity increased, theory of mind followed (Cohort 2). The executive factor fits the universality of these sequences well.

### IV.IV. Conceptual competence on the line: Squaring the executive factor with early competence on nonverbal tests of theory of mind

Emerging research suggests that infants pass theory of mind tasks that are nonverbal (Onishi & Baillargeon, 2005; Surian, Caldi & Sperber, 2007; Southgate, Senju & Csibra, 2007). In these studies, infants witness a nonverbal unexpected transfer event, and then automatically (and correctly) anticipate where the protagonist of the story will search for the object by orienting their gaze in the direction of the original location of the object (the false belief location). Nonetheless, young preschoolers fail our verbal measures of false belief attribution consistently, before they begin to consider multiple responses and finally enter a period of consistent passing.

What performance factors limit a child's ability to express their competence in a verbal paradigm, when we know they are capable of understanding the concepts in a nonverbal paradigm? Executive functions may be the limiting factor in verbal paradigms where children must deliberately respond to a question about someone's belief. When the child must give a deliberate answer, they are required to consider multiple response options and select one of them, then produce this answer either verbally or by pointing. In nonverbal paradigms, there is no such requirement to deliberately select and then produce an answer. Children's gaze indicates understanding that is automatic. The hypothesis is that automatic operation of an innate mechanism taps less (or different) executive resources than the deliberate operation of the very same mechanism.

Of course, the executive resources which make verbal tasks more difficult than nonverbal ones are likely to be different than the executive resources which make false belief tasks more difficult than true belief tasks. I raise the issue to point out that the executive components may enter into cognitive development in more than one way.

### IV.IV.1. Executive functions and domain specificity of theory of mind

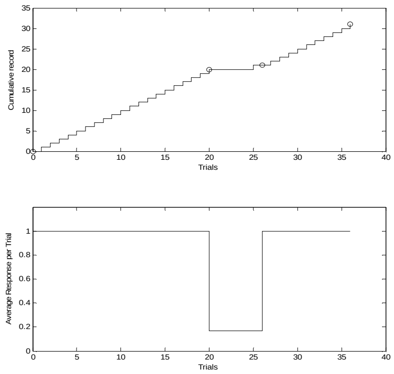
What this implies about our theories of developing theory of mind is that change could be quantitative, rather than qualitative. In other words, differences in children's capacity to reason about mental states, as measured by their gaze direction or their verbal answers in our false belief paradigms, may arise because different response modalities are engaged. In theory of mind reasoning, when gaze is the indicator of conceptual competence, children show earlier competence than when a verbal answer is required. Whether this phenomenon can be generalized to other areas of conceptual competence is an empirical question, though there is evidence to suggest it can. It is not unheard of for different answers to be ascribed to children depending upon the modality of response we measure. For example, Hood, Cole-Davies & Dias (2003) found that two- and three-year-olds's gaze and manual search for a hidden object rolling down a ramp were not correlated. Measuring the children's gaze direction generally yielded a higher rate of correct answers in young children than did a measure of their choice in a manual search. Here too, a crucial difference between these tasks and the response modalities accompanying them is the degree to which a response must be deliberated before it is given.

### IV.IV.2 Applications

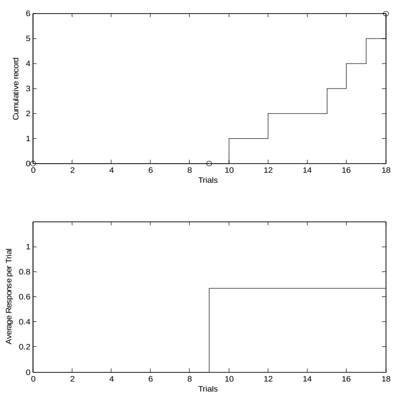
Detecting a mismatch in modalities can be useful in education. Church & Goldin-Meadow (1986) found that some children's gestures were not consistent with their verbal explanations of a Piagetian conservation task. The authors contrasted children whose gesture and speech were mismatched with children whose gesture and speech were consistent. Using a measure of internal consistency of the verbal explanations, they found that knowledge was less internally consistent in the mismatch group than in the match group, suggesting that a mismatch in response modalities is an indicator of conceptual instability. They also found that the "mismatched" children were more sensitive to instruction than the "matched" children. With replications and extensions of these findings across domains of cognitive development we could hope to see early indices of conceptual competence, and make these means for detecting early competence available to clinicians and educators alike. The period of instability in the expression of conceptual understanding documented by our microgenetic studies may be similar to what Church & Goldin-Meadow (1986) and Hood, Cole-Davies & Dias (2003) reported. Perhaps it corresponds to Vygotsky's zone of proximal development when the child is most receptive to outside intervention (Vygotsky, 1935). It would be worth testing this prospective developmental hypothesis empirically, using training or transfer designs (see Kuhn, 1995, and Amsterlaw & Wellman, 2006, for examples). If this is the case, it would be desirable for a teacher to notice when a pupil's answers enter a phase of instability.

Our findings within individual's developmental records and others' findings support the suggestion that expressing early competence in mental state reasoning is subject to developing executive functions. Whether and how well this theory will generalize across domains of development (e.g., naïve physics or mathematics) is a question for further investigation (see Hood, Wilson & Dyson, 2006). How is the individual child's performance in gravity tasks related to developing executive functioning? Do we observe the same types of intra-individual instability, along with inter-individual variability in acquisition curves?

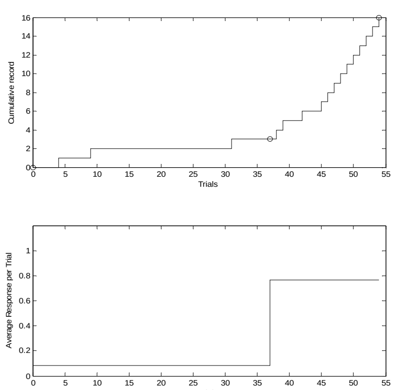
Figure 5. Cumulative records for S1.




S1 True Belief performance

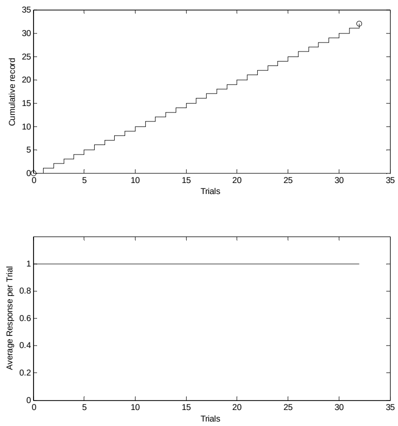



S1 False Belief look first performance

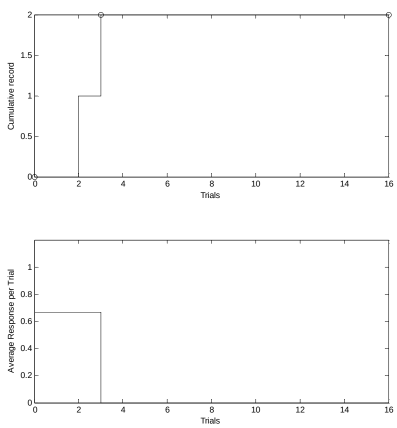



S1 Classic False Belief performance

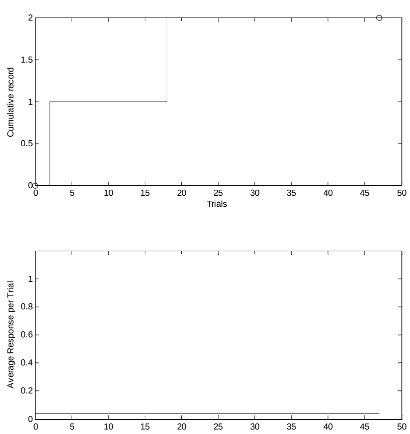
Figure 6. Cumulative records for S2.




S2 True Belief performance

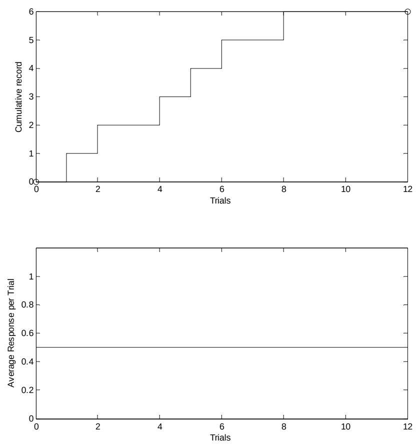



S2 False Belief look first performance

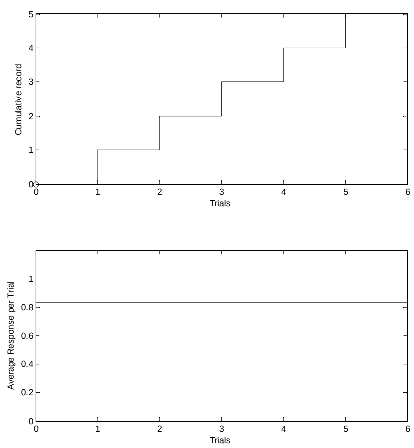



S2 Classic False Belief performance

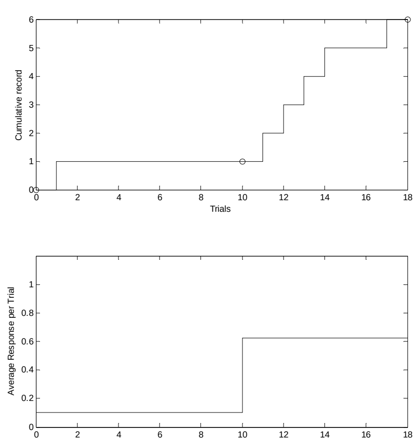
Figure 7. Cumulative records for S3.




S3 True Belief performance

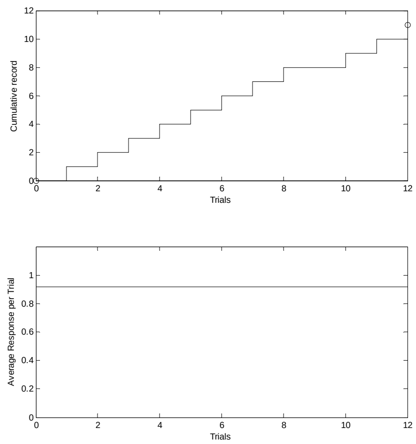



S3 False Belief look first peformance

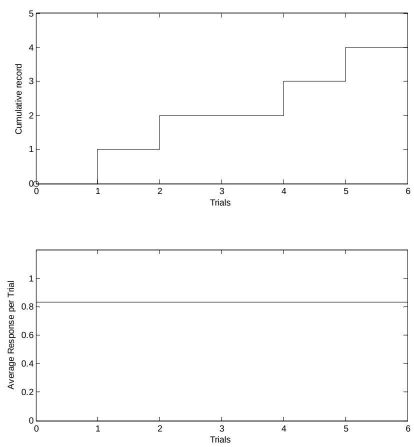



S3 Classic False Belief Performance

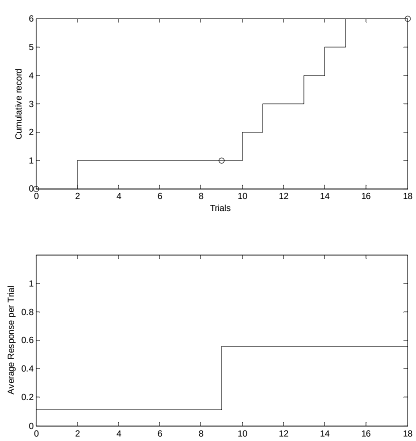
Figure 8. Cumulative records for S4.




S4 True Belief performance

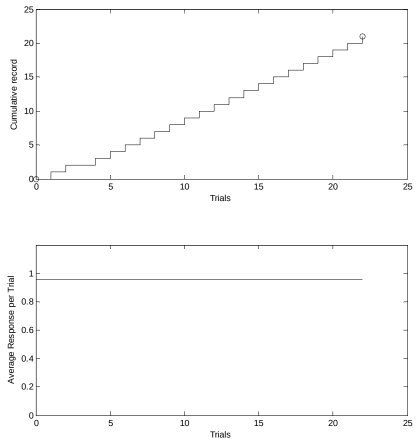



S4 False Belief look first performance

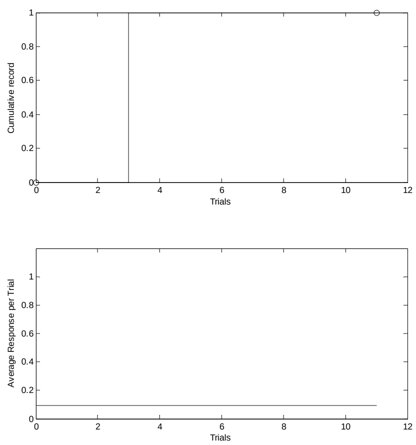



S4 Classic False Belief performance

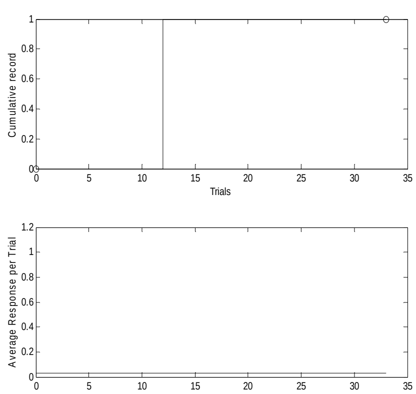
Figure 9. Cumulative records for S5.




S5 True Belief performance

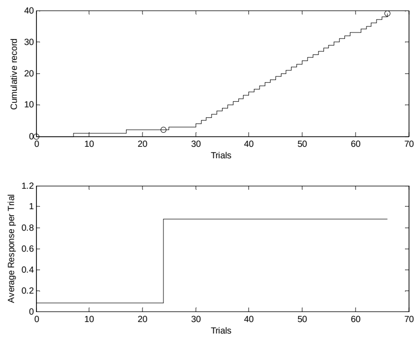



S5 False Belief look first performance

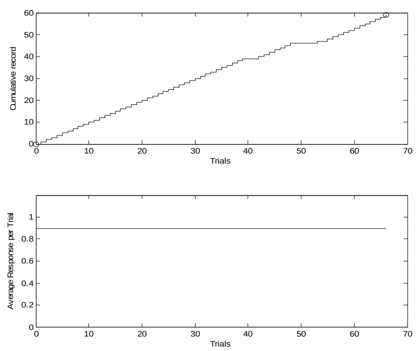



S5 Classic False Belief performance

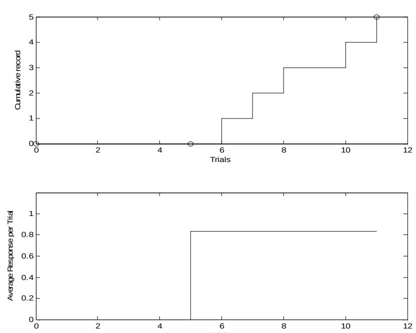
Figure 10. Cumulative records for S6.



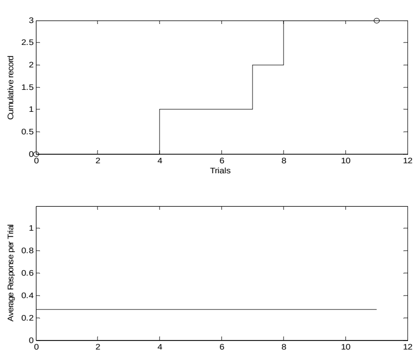

S6 Incongruent performance




S6 Congruent performance

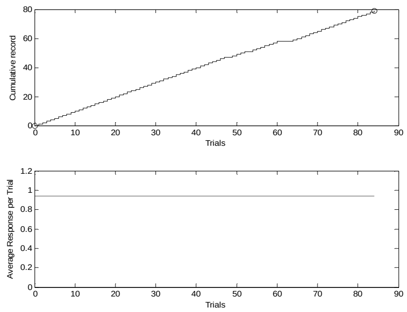



S6 False Belief unexpected transfer performance

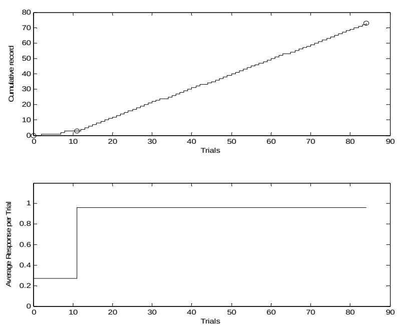



S6 False Belief unexpected contents performance

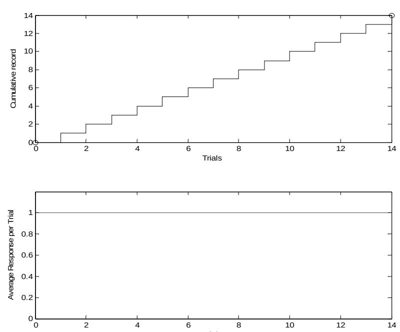
Figure 11. Cumulative records for S7.



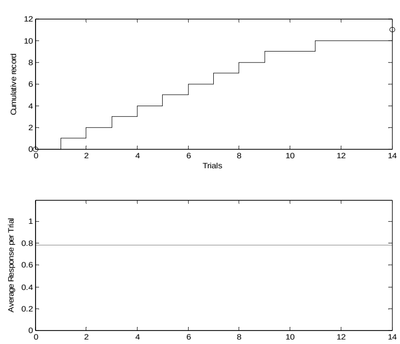

S7 Incongruent performance




S7 Congruent performance

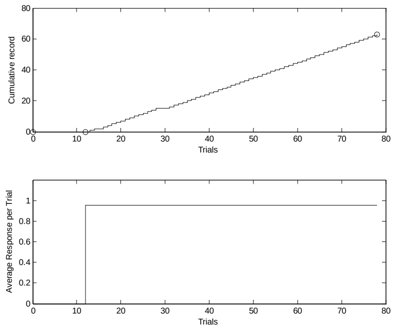



S7 False Belief unexpected transfer performance

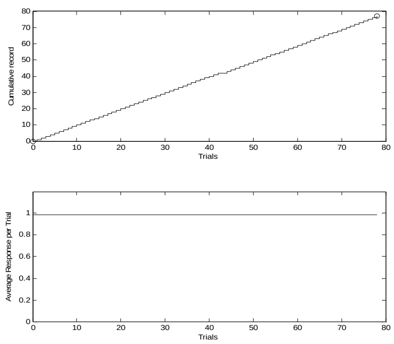



S7 False Belief unexpected contents performance

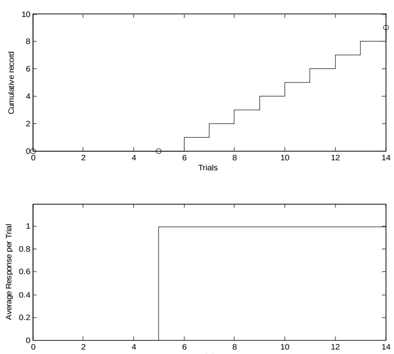
Figure 12. Cumulative records for S8.



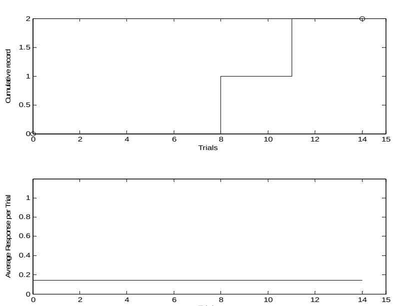

S8 Incongruent performance




S8 Congruent performance

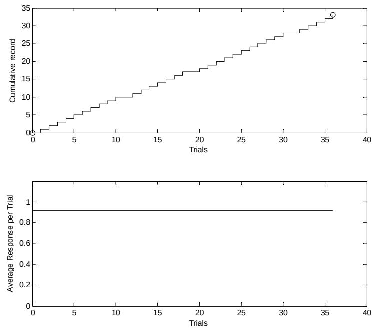



S8 False Belief unexpected transfer performance

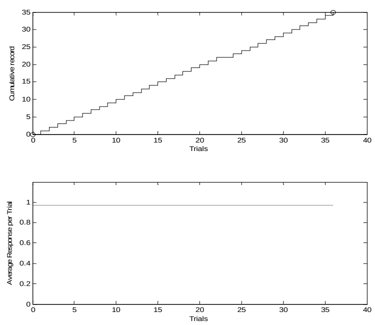



S8 False Belief unexpected contents performance

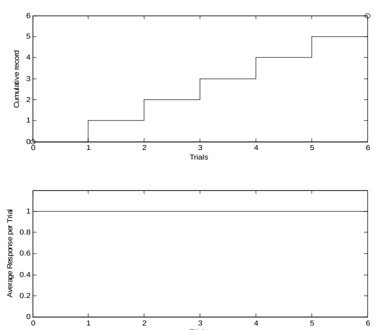
Figure 13. Cumulative records for S9.



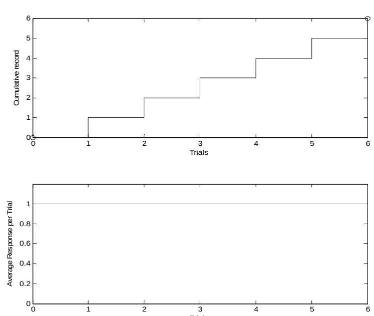

S9 Incongruent performance




S9 Congruent performance

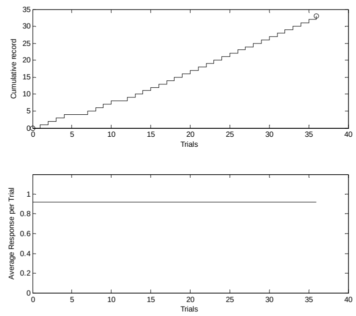



S9 False Belief unexpected transfer performance

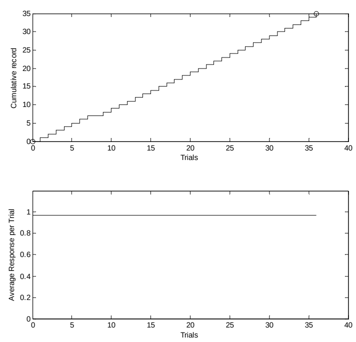



S9 False Belief unexpected contents performance

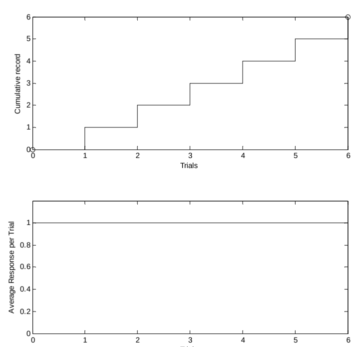
Figure 14. Cumulative records for S10.



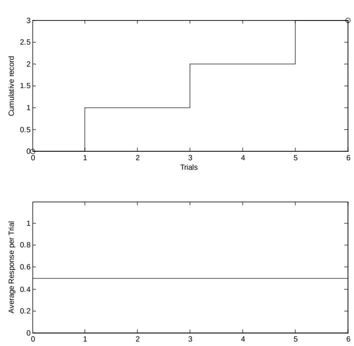

S10 Incongruent performance




S10 Congruent performance

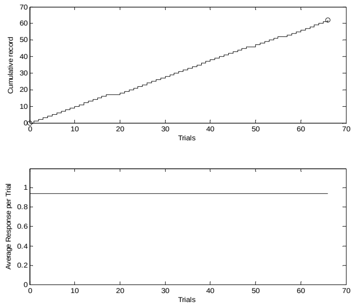



S10 False Belief unexpected transfer performance

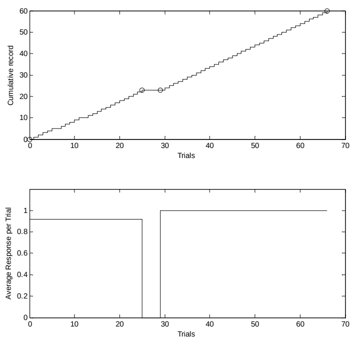



S10 False Belief unexpected contents performance

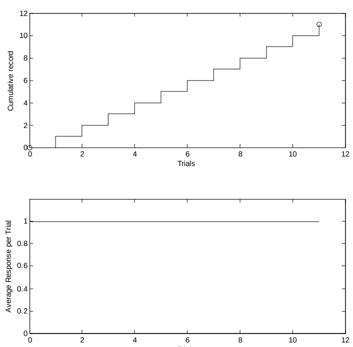
Figure 15. Cumulative records for S11.



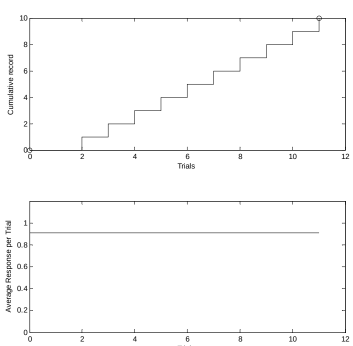

S11 Incongruent performance




S11 Congruent performance




S11 False Belief unexpected transfer performance




S11 False Belief unexpected contents performance

Appendix 1

Example session for cohort 1

Task order and question order was counterbalanced across sessions. Also, the direction of transfer of the object, left to right on the stage or right to left on the stage, was counterbalanced. Sometimes three locations were available as answer options and sometimes two locations were available.

False belief- unexpected transfer

Steve and Corrie are playing here in their room. In their room there is a basket and a toybox. Now Steve is going for a walk. He leaves his ball here in the basket and goes outside. He can't see us anymore. While Steve is outside, look what happens! Corrie takes the ball from here, and he puts it in here. Now Steve didn't see that happen, did he? No. Well, Steve is coming back soon, and I have some questions for you.

Where will Steve look for the ball? (test question)

Where is the ball now? (control reality question)

Where did Steve put the ball, in the beginning of the story? (control memory question)

Where does Steve think the ball is? (test question)

False belief- unexpected transfer look first

Here's Molly and Mike. They have a new house. Look, there's a barrel, a cup and a cupboard. Molly has a lamp to put in her house. She decides to put it in the cupboard for now, and goes to get more furniture. She can't see us because she left, far away. Now Mike takes the lamp and puts it in the barrel. Molly is coming back and I want to ask you something.

Where does Molly think the lamp is? (test question)

Where did she put the lamp in the beginning of my story? (control memory question)

Where is the lamp right now? (control reality question)

Where will Molly look first for the lamp? (test question)

True belief- expected transfer

This is a school house with some kids inside. They are having class. Today, Dorris brought her friend dinosaur to class. And there are some boxes in the classroom too. See – a red box, a blue box and a green box. Dorris needs to clean her glasses so she goes outside to get some soap. She leaves her glasses in the red box. When she comes back with the soap she sees the dinosaur take the glasses and put them under the green box. She saw him do it, because she was standing right there watching him, right? Now let me ask you something.

Where does Molly think her glasses are? (test question)

Where are her glasses really? (control reality question)

But which box did she leave the glasses in, when she went outside? (control memory question)

Where will Molly look for her glasses? (test question)

Appendix 2

Example session for cohort 2. Props were varied across sessions. Task order and question order was counterbalanced across sessions.

<u>False belief- unexpected transfer</u>

The direction of transfer of the object, left to right on the stage or right to left on the stage, was counterbalanced.

"Steve and Corrie are playing here in their room. In their room there is a basket and a toybox. Now Steve is going for a walk. He leaves his ball here in the basket and goes outside. He can't see us anymore. While Steve is outside, look what happens! Corrie takes the ball from here, and he puts it in here. Now Steve didn't see that happen, did he? No. Well, Steve is coming back soon, and I have some questions for you."

"Where is the ball now?" (control reality question)

"Where did Steve put the ball, in the beginning of the story?" (control memory question)

"Where does Steve think the ball is?" (test question)

<u>False belief- unexpected contents</u>

"I have something I want to show you."

(Show a crayon box).

"What do you think is inside?" (control question)

(Open box to reveal unexpected contents – a sock.)

“Now, if we show this box to Billy in your class, what will Billy think is inside?”

(test question)

Footnotes

(1) I deliberately chose not to discuss a third competing view, the simulationist account of mental state reasoning.
(2) The original paper presenting this task suggested that this was an analog to the Stroop task for children (Gerstadt, Hong & Diamond, 1994). Note that the 'Day/ Night' condition is like Stroop's experiment 2, and the 'Dog/ Pig' condition is like Stroop's experiment 1.
(3) No child participated in more than one experiment reported here.
(4) The alpha level is set manually when implementing the algorithm. In my analyses I tested for change points with various alpha levels, up to alpha = .1. All results reported here are for alpha < .05. Though the cut-off is arbitrary, we found no change points for an alpha greater than .05 which were not already present at lower levels. See Papachristos & Gallistel (2006) for a discussion.

References

Amsterlaw, J., & Wellman, H.M. (2006). Theories of mind in transition: A Microgenetic study of the development of false belief understanding. *Journal of Cognition and Development*, *7*, 139-172.

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

Bartsch, K. & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, *60*, 946-964.

Carlson, S.M., Mandell, D.J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from age 2 to 3. *Developmental Psychology,* 1105-1122.

Carlson, S.M., & Moses, L.J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032-1053.

Carlson, S.M., Moses, L.J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibition and working memory. *Infant and Child Development*, *11*, 73-92.

Carlson, S.M., Moses, L.J., & Claxton, L.J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, 299-319.

Clements, W., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377- 397.

Church, R.B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*, 43-71.

Denckla, M.B., & Reader, M.J. (1993). Education and psychosocial interventions: Executive dysfunction and its consequences. In R. Kurlan (Ed.), Handbook of Tourette's Syndrome and Related Tic and Behavioral Disorders. New York: Marcel Dekker, 431-451.

Diamond, A., Kirkham, N.Z., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, **38**, 352-362.

Dickson, C.A., Deutsch, C.K., Wang, S.S., & Dube, W.V. (1996). Matching-to-sample assessment of stimulus overselectivity in students with intellectual disabilities. *American Journal on Mental Retardation*, 447-453.

Dube, W.V., McIlvane, W.J., & Green, G. (1992). An analysis of generalized identity matching-to-sample test procedures. *Psychological Record*, 17-28.

Dunn, & Dunn (1981). Peabody Picture Vocabulary Test- Revised.

Espy, K.A, McDiarmid, M.M., Cwik, M.F., Stalets, M.M., Hamby, A., & Senn, T. (2004).The Contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, 465-486.

Flavell, J.H., Green, F.L., & Flavell, E.R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development*, *51*.

Flavell, J.H., & Wohlwill, J.F. (1969). Formal and functional aspects of cognitive development. In *Studies in Cognitive Development: Essays in Honor of Jean Piaget*. D. Elkind & J. Flavell (Eds). New York: Oxford University Press.

Flynn, E. (2006). A Microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology*, 631- 654.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, 7, 103-115.

Friedman, N.P., & Miyake, A. (2004). The Relations among inhibition and interference control functions: A Latent-variable analysis. *Journal of Experimental Psychology: General*. 101-135.

Friedman, O., & Leslie, A.M. (2004a). Mechanisms of belief-desire reasoning:Inhibition and selection. *Psychological Science*, *15*, 547-553.

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

Friedman, O., & Leslie, A.M. (2005). Processing demands in belief-desire reasoning: Inhibition or general difficulty? *Developmental Science*, *8*, 218-225.

Friedman, O., Baker, S.T., & Leslie, A.M. (submitted). Does Belief-Desire Reasoning Use Different Inhibitory Processes than 'Conflict' Inhibition?

Frye, D., Zelazo, P.D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*, 483-527.

Gallistel, C.R., Fairhurst, S, & Balsam, P. (2004). The Learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Science*, *101*, 13124-13131.

Gelman, R. (2000). Domain specificity and variability in cognitive development. *Child Development*, *71*, 854-856.

Gelman, R., & Williams, E. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. Siegler (Eds), N. Damen (Series Ed), *Handbook of Child Psychology*: *Vol. 2. Cognition, perception and language*. (5th ed., pp. 575-630). New York: Wiley.

Gerstadt, C., Hong, Y., & Diamond, A. (1994). The relationship between cognition and action: Performance of 3 ½ -7 year old children on a Stroop-like day-night test. *Cognition*, *53*, 129-153.

Gopnik, A. (1993). How we know our own minds: The Illusion of first person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*, 1-14.

Gopnik, A., & Astington, J. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26-37.

Hogrefe, G., J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in the attribution of epistemic states. *Child Development*, *57*, 567-582.

Hood, B.M., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology*, *39*, 61-70.

Hood, B.M., Wilson, A., & Dyson, S. (2006). The Effect of divided attention on inhibiting the gravity error. *Developmental Science*, 303-308.

Hughes, C. (1998). Finding your marbles: Does Preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology*, 34, 1326- 1339.

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

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. *Developmental Psychology, 29,* 498-510.

Kochanska, G., Murray, K., Jacques, T.Y., Koenig, A.L., & Vandegeest, K.A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*, 490-507.
Leslie, A.M. (1987). Pretense and representation: The Origins of theory of mind. *Psychological Review*, *94*, 412-426.
Leslie, A.M. (2000). How to acquire a 'representational theory of mind'. In D. Sperber (Ed.). *Metarepresentations: A Multidisciplinary Perspective*. (pp. 197-223). Oxford: Oxford University Press.
Leslie, A.M., German, T., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, *50*, 45-85.
Leslie, A., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*, 247-253.
Leslie, A.M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43, 225–251.
Luciana, M, Conklin, H.M., Hooper, C.J., & Yarger, R.S. (2005). The Development of nonverbal working memory and executive control processes in adolescents. *Child Development*, 697-712.
Luria, A.R. (1973). The Working Brain: An Introduction to Neuropsychology. New York: Basic Books.
MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation*, Vol. 43 (pp. 163-214). San Diego, CA: Academic Press.
MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An Integrative review. *Psychological Bulletin*, *109*, 163-203.
Mahone, E.M., Koth, C.W., Cutting, L., Singer, H.S., & Denckla, M.B. (2001). Executive function in fluency and recall measures among children with Tourette syndrome or ADHD. *Journal of the International Neuropsychological Society*, *7*, 102-111.
Meltzer, L., Katzir-Cohen, T., Miller, L., & Roditi, B. (2001). The Impact of effort and strategy use on academic performance: Student and teacher perceptions. *Learning Disability Quarterly*, 85-98.
Moore, C., Pure, K., & Furrow, P. (1990). Children's understanding of the modal expression of certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*, 722- 730.
Noveck, I.A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165-188.
Onishi, K., & Baillargeon, R. (2005) Do 15-Month-Old Infants Understand False Beliefs? *Science*, 255-258.
Ozonoff, S, Pennington, B.F., & Rogers, S.J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship with theory of mind. *Journal of Child Psychology and Psychiatry*, *32*, 1081-1105.
Papachristos, S., & Gallistel, C.R. (2006). Autoshaped head poking in the mouse: A Quantitative analysis of learning curve. *Journal of the Experimental Analysis of Behavior*, *85*, 293-308.
Pears, K.C., Moses, L.J. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development*, *12*, 1-20.
Perner, J., Frith, U., Leslie, A.M., & Leekam, S.R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Development*, *60*, 689-700.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*, 337- 344.
Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development*, *73*, 752- 767.
Perner, J, Leekam, S., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 125-137.
Rein, J. (2005). Using individual developmental curves to study false belief attribution in preschoolers. *Unpublished thesis.*
Roth, D, & Leslie, A.M. (1991). The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Developmental Psychology*. 315-330.
Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, *9*, 331-349.
Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, 17, 74-81.
Schulz, L., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*, 162-176.
Senn, T.E., Espy, K.A., & Kaufmann, P.M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology*, 445-464.
Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*, 1-12.
Siegler, R.S. (1996). *Emerging minds: The Process of change in children's thinking*. New York: Oxford University Press.
Siegler, R.S. (2007). Cognitive variability. *Developmental Science*, *10*, 104-109.
Siegler, R.S., & Svetina, M. (2002). Microgenetic/ Cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development*, *73*, 793-809.
Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*, 587-592.
Surian, L., & Leslie, A.M. (1999). Competence and performance in false belief understanding: A comparison of autistic and three-year-old children. *British Journal of Developmental Psychology*, 17, 141–155.
Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*, 580-586.
Vygotsky, L.S. (1962). *Thought and Language*. Cambridge: MIT Press.
Wellman, H. (1988). First steps in the child's theorizing about the mind. In *Developing Theories of Mind*. Astington, J.W., Harris, P.L., & Olson, D.R. (Eds). Cambridge: Cambridge University Press.
Wellman, H. (1998). Culture, variation, and levels of analysis in folk psychologies: Comment on Lillard (1998). *Psychological Bulletin*, *123*, 33-36.
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.

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. (1991). Is only seeing really believing? Sources of true belief in the false belief task. *Cognitive Development*, *6*, 91-103.

Zelazo, P.D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37-63.

Zelazo, P.D., Mueller, U., Frye, D., & Marcovitch, S. (2003). The Development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, *68*.

## *Curriculum Vita*

### Sara Theresa Baker, MS

2004 **Master of Science**, Department of Psychology, Rutgers University
Advisor: Dr. Alan Leslie
Thesis: "Do opposites attract? A new test of inhibitory processing in preschoolers"

2002 **Maîtrise**, General experimental psychology, Honors "très bien"
Université Paris 8, France; Advisor: Dr. Guy Politzer
Thesis: "Revision of conditional beliefs by 6- and 7-year-old children"

2001 **Licence**, Human sciences, Social psychology, Honors "assez bien"
Université Paris 8, France
Thesis (literature review): "Public Opinions"

2000 **DEUG**, Human sciences, Psychology, Université Paris 8, France

1996 **Baccalauréat**, Literature, Languages and Philosophy, Honors "assez bien"
Lycée Français de Vienne, Austria