

Relations Between Theory of Mind and Executive Function in Middle Childhood: a Short-Term Longitudinal Study

Abstract

Studies with preschool children have shown significant links between children's executive function (EF) and theory of mind (ToM), but few studies have examined these associations in primary school children. In order to address this gap we designed a three-wave cross-lagged longitudinal study in which we followed a group of 113 children (61 boys) across three time points from age 9.5 to 10.5 years (mean age at time 1 $M = 112.3$ months, $SD = 4.18$ months; mean age at Time 2 $M = 118.3$ months, $SD = 4.15$ months and mean age at Time 3 $M = 124.7$ months, $SD = 4.06$ months). At each time point we measured EF (working memory and inhibitory control), ToM and language. Our analyses showed (i) moderate rank-order stability of individual differences in both EF and ToM and (ii) growth in ToM task performance across time. Cross-lagged longitudinal analyses revealed an asymmetric developmental relation between ToM and working memory. Early working memory predicted later ToM but not vice versa. Our results suggest a specific role for working memory in the on-going development of ToM in middle childhood.

Keywords: Theory of Mind, Executive Function, Middle Childhood

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The final article is available at:

<https://doi.org/10.1016/j.jecp.2017.06.011>

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Relations Between Theory of Mind and Executive Function in Middle Childhood: a Short-Term Longitudinal Study

The ability to infer others' mental states or 'theory of mind' (ToM) has long fascinated researchers from a wide range of disciplines. Individual differences in children's ToM are striking and related to important social (Banerjee, Watling, & Caputi, 2011) and cognitive outcomes (Lecce, Caputi, & Hughes, 2011). Specifically, children who excel in reasoning about others' mental states are better equipped to face the demands of formal schooling, as they are more academically able (Blair & Razza, 2007; Lecce, Caputi, & Pagnin, 2014), have greater metacognitive knowledge (Lecce, Bianco, Demicheli, & Cavallini, 2014; Lecce, Zocchi, Pagnin, Palladino, & Taumoepeau, 2010), and are more socially competent (Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016). To date, most of the progress made on understanding ToM has concerned dramatic developments in mindreading in the preschool years (i.e., false belief understanding). While interest in infants' and adults' ToM has grown, the study of middle childhood is more recent (Hughes & Devine, 2015a). This expansion of research on ToM into middle childhood raises important questions about the nature of developmental changes in ToM and associations with age-related improvements in domain-general cognitive skills. To address these issues, the present study adopted a short-term longitudinal design. This approach has merits from both theoretical and practical perspectives. Theoretically, our short-term longitudinal design enables us to test the role of EF in the development and expression of ToM in middle childhood and to examine whether results found in preschoolers can be extended to older children (see sections below). At a practical level, we hoped that our findings would yield insights on how to promote ToM development.

Explaining the Relations Between Theory of Mind and Executive Function

Numerous studies have shown that individual differences in executive function (EF) independently predict variation in ToM in early childhood (Devine & Hughes, 2014). EF is an umbrella term describing the higher-order processes for controlling, directing and monitoring cognitive function (Diamond, 2013; Zelazo, Carlson, & Kesek, 2008). EF, thus, encompasses a rather heterogeneous collection of skills that include inhibitory control (IC), working memory (WM), set-shifting and planning (Miyake & Friedman, 2012).

Correlational studies have consistently shown moderate associations between individual differences in EF and ToM, even when the effects of age and verbal ability are controlled (Carlson & Moses, 2001; Hughes, 1998). In a meta-analysis drawing together data from nearly 10,000 children aged between 3 and 6 years, Devine and Hughes (2014) reported a moderate association between EF and false belief understanding that was similar for children belonging to different cultures and consistent across distinct EF tasks.

One interpretation of these results is that the relation between EF and false belief understanding is peripheral and arises from the incidental demands that standard false-belief tasks place on EF (e.g., Perner, Lang, & Kloo, 2002). According to this *expression* account, children need EF to display their ToM understanding. At least four strands of evidence challenge the expression account. First, meta-analysis demonstrates significant correlations between early EF and later ToM, even when initial ToM scores are taken into account (Devine & Hughes, 2014). Second, cross-cultural research shows that cultural groups (e.g., Chinese children) who show higher levels of EF relative to North American or British children, do not exhibit a parallel advantage in ToM (Wang, Devine, Wong & Hughes, 2016). Third, individual differences in EF remain correlated with performance on false-belief tasks in which executive demands are reduced (Henning, Spinath, & Aschersleben, 2011). Fourth, EF shows similarly strong associations with first- and second-order false belief tasks, even though the latter place greater demands on EF (Devine & Hughes, 2014). These four strands

of evidence support the view from the '*emergence*' account (Russell, 1996) that EF contributes to the *development* of ToM and suggest that there is a deep and unique relationship between ToM and EF that cannot be explained by only considering the executive demands of ToM tasks.

Since most of what is known about the relations between EF and ToM is based on data from early childhood, the next step for research is to examine these relations in middle childhood and beyond. Adopting the principle of parsimony, we predicted that, as in the preschool years, EF would contribute to the *emergence* rather than *expression* of ToM in middle childhood.

Development of Theory of Mind and Executive Function in Middle Childhood

The dearth of research on the links between ToM and EF in middle childhood is surprising given that important age-related improvements across middle childhood have been reported for both ToM (Devine & Hughes, 2013) and EF (Davidson, Amso, Anderson, & Diamond, 2006; Huizinga, Dolan, & Van Der Molen, 2006). These developmental changes in ToM and EF are relevant as they contribute to children's social and cognitive functioning in this period of life. Specifically, ToM understanding independently predicts children's ability to: i) comprehend teachers' comments (Lecce et al., 2011; Lecce, Caputi, et al., 2014); ii) establish good social relations (Slaugther et al., 2015); and iii) show better social competence (Devine et al., 2016) and academic achievement (Lecce et al., 2014). Regarding EF, Jacobson and colleagues showed that EF in late primary school uniquely predicted between 3% and 13.5% of the variance in children's academic and behavioral functioning in Year 6 (Jacobson, Williford, & Pianta, 2011). Moreover, regulatory skills independently contributed to predict variance on social functioning in 10- to 12-year-olds (Murphy, Shepard, Eisenberg, & Fabes, 2004).

In the present study, we focused on children aged between 9 and 10 years. The few available studies of ToM in middle childhood clearly show developmental changes across primary school (Banerjee et al., 2011; Devine & Hughes, 2013; Rice, Anderson, Velnoskey, Thompson, & Redcay, 2016). For example, 11-year-old children are more skilled than children aged 9 in understanding faux pas (Baron-Cohen, O’Riordan, Stone, Jones, & Plaisted, 1999) and Im-Bolter, Agostino and Owens-Jaffray (2016) reported significant differences between 7- to 8-year-old children and 11- to 12-year-old children on the Strange Stories task.

Exactly how these observed age-related differences in performance on a range of measures of ToM should be interpreted remains open to debate. One possibility is that age-related differences in performance on ToM tasks in middle childhood could be explained by age-related gains in verbal ability or general cognitive ability. Several studies now show that age makes a unique contribution to individual differences in performance on measures of ToM even controlling for verbal ability (Devine & Hughes, 2013; 2016).

A second possibility is that children undergo further meta-representational changes in their ability to represent another’s belief (see Apperly, 2011 and Miller, 2009). This account holds that there are further breakthroughs and discoveries to be made about the nature of mental-state representation beyond the one required to pass false belief tasks. According to Apperly (2011), this account struggles on several fronts. It is difficult to identify what new insights in the ability to represent another’s belief would be required to navigate mental life beyond an understanding of desires, knowledge and beliefs. In the Strange Stories task, for example, children are asked to apply their understanding of a character’s beliefs, knowledge or desires to explain a character’s actions. Despite simply requiring an understanding of these mental states, there is considerable evidence for continued age-related gains in performance on this task across middle childhood (Devine & Hughes, 2013; 2016).

Moreover, similar tasks, such as the Faux Pas task, only require participants to make distinctions between their own and the character's beliefs or knowledge. Despite relying on a participant's false belief understanding, there is evidence for within-person gains in performance on this task across middle childhood (Banerjee et al., 2011). If the ability to represent another's belief was all that mattered then one would not expect age-related gains in performance on either of these tasks.

A third possibility is that children's ability to *use* their understanding of mental states in a more flexible and appropriate way improves with age (Apperly, 2011; Bianco, Lecce, & Banerjee, 2016; Devine & Hughes, 2013; Hughes & Devine, 2015a). Over middle childhood children develop the ability to simultaneously infer a variety of mental states in complex social scenarios and develop a more complex awareness of the links among multiple mental states (Amsterlaw, Lagattuta, & Meltzoff, 2009; Bender, Pons, Harris, & De Rosnay, 2011; Lagattuta, 2014). For example, in the Faux Pas task, children are required to understand simultaneously the intentions, the knowledge state and the emotions of the characters. Interestingly, if we look at the single questions that follow each Faux Pas story it is clear that all typically developing primary school children should be able to make each of the above inferences, when considered separately, as we know that by age 5 children should be able to attribute ignorance to others and to understand external causes of simple positive/negative emotions and intentions (Pons, Harris, & De Rosnay, 2004; Wellman & Liu, 2004). What seems to be challenging, therefore, is not the content of each of the mental inferences, but the combined focus on knowledge states, intentions, and emotions and the links between them. This ability to align emotions, decisions and thoughts, when referred to a single social event, shows a clear developmental trajectory over middle childhood (Lagattuta, Elrod, & Kramer, 2016). Similar conclusions can be drawn with respect to the Strange Stories and the Silent Film task for which a clear progression through middle childhood has been found (Devine &

Hughes, 2016). In addition, these two tasks require children to apply their ToM skills not in a single, although complex, situation (like in the Faux Pas) but in a variety of contexts (Miller, 2012). This variety is challenging for children, as they are required to make mental states inferences that are appropriate to capture the salient aspect of the given social situation (Miller, 2012). There is, therefore, a difference in ToM skills between a child who is able to infer a plausible (and thus appropriate) mental state in a social situation and a child who is able to infer the right mental state that grasps the focal point of what is happening (i.e., double bluff, misunderstanding and so on) (Bianco et al., 2016).

Regarding EF, perhaps one of the most notable changes across middle childhood concerns the structure of EF itself. EF has been conceptualized as a multi-faceted construct made up of three distinct but closely related components: WM, shifting and inhibition (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Developmental research suggests that these components are robustly inter-correlated in preschoolers and tend to fractionate as children progress through middle childhood. Indeed, whereas a single latent factor seems sufficient to account for preschoolers' performance (Hughes, Ensor, & Wilson, 2010; Wiebe, Espy, & Charak, 2008), this is not the case in middle childhood and adolescence (Miyake & Friedman, 2012; McAuley & White, 2011; van der Sluis, de Jong, & van der Leij, 2007). Overall, the structure of EF gradually separates with age, developing from a cohesive, general ability to more distinct, specific abilities (Shing, Lindenberger, Diamond, Li, & Davidson, 2010; Xu et al., 2013). These results provide support for the non-unitary nature of EF in middle childhood and adolescence and highlight the need to consider EF components separately when investigating older children. Drawing on data from three time-points in middle childhood, the first aim of our study was to examine the development and rank-order stability of ToM and EF.

Relations Between Theory of Mind and Executive Function in Middle Childhood

Recent years have seen a growing interest in the relations between ToM and EF beyond the preschool years. Cross-sectional studies indicate that the relation between performance on ToM tasks and measures of EF is weak to moderate in adolescence (Vetter, Altgassen, Phillips, Mahy, & Kliegel, 2013) and middle childhood (Bock, Gallaway, & Hund, 2015; Cantin, Gnaedinger, Gallaway, Hesson-McInnis, & Hund, 2016; Lagattuta, Sayfan, & Harvey, 2014; Wang et al., 2016). However, relatively few studies (e.g., Carlson, Moses, & Breton, 2002) have examined the specific links between distinct aspects of EF and children's ToM. While the meta-analytic evidence based on preschool children suggests that there is a consistent link between distinct measures of EF and false-belief task performance (Devine & Hughes, 2014), it is entirely possible that the relations between ToM and specific aspects of EF may shift as EF becomes more fractionated. Addressing this possibility, our second aim was to examine the links between individual differences in ToM and in two key components of EF: WM and IC. We chose to focus on WM and IC as they appear to have the strongest association with ToM during early childhood (Devine & Hughes, 2014; Marcovitch et al., 2015). We did not include a measure of cognitive flexibility to minimize the burden on participants and because existing studies indicates that cognitive flexibility plays a relatively minor role in more mature ToM (Ahmed & Miller, 2011, Wang et al., 2016).

With few exceptions (Austin, Groppe, & Elsner, 2014; Devine et al., 2016), the great majority of existing studies on EF in middle childhood are cross-sectional. Unlike cross-sectional studies, longitudinal designs, and more specifically cross-lagged longitudinal designs, provide a platform for testing developmental theories about the relations between different constructs. While cross-sectional studies can reveal whether or not two variables are related at a specific point in time, cross-lagged longitudinal designs are needed to determine whether one variable is implicated in the development of another variable

(Newsom, 2015). Cross-sectional findings from middle childhood provide some evidence for an ongoing relation between ToM and EF beyond the preschool years, but do not enable any further analysis of the *developmental* relations between these constructs. Cross-lagged longitudinal research provides one way to begin to unpack the association between ToM and EF beyond the preschool years.

In the first study of its kind, Austin et al. (2014) examined the relations between a composite measure of ToM and three EF domains: attention shifting (Cognitive Attention shifting task, Röthlisberger, Neuenschwander, Michel, & Roebbers, 2010), WM (Digit-Span Backward task, Petermann and Petermann, 2007) and IC (Fruit Stroop task, Archibald & Kerns, 1999). ToM was measured by summing children's scores on two second-order false-belief tasks (Perner & Wimmer, 1985), the Strange Stories (Happé, 1994), the Extended Theory-of-Mind Scale (Henning, Hofer, & Aschersleben, 2012) and a cartoon sequencing ToM task (Völlm et al., 2006). This study, involving 1657 children aged between 6 and 11 years followed across a 12-month period, showed weak concurrent associations between each measure of EF and ToM task performance and very weak (but statistically significant) cross-lagged associations between each indicator of EF and ToM performance (all β s < .10). Despite these very weak effects, the authors concluded that there was a reciprocal relation between EF and ToM. However, at both time points the sample varied widely in age (i.e., a range of 5 years), which coupled with the lack of a control measure of verbal ability makes these findings even more difficult to interpret.

In another longitudinal study Devine et al. (2016) examined the links between composite measures of ToM and EF at two time points spanning a 4-year period in middle childhood. ToM was measured via classical first- and second-order false-belief tasks at Wave 1 and via the Strange Stories (White, Hill, Happé, & Frith, 2009), the Silent Film (Devine & Hughes, 2013) and the Triangles (Castelli, Happé, Frith, & Frith, 2000) tasks at

Wave 2. EF was assessed using a conflict-inhibition task (Gerstadt, Hong, & Diamond, 1994) together with a shifting task (Trucks Game, Hughes & Ensor, 2005), a planning task (Tower of London, Shallice, 1982), and a WM task (Bead memory task, Thorndike, Hagen, & Sattler, 1986) at Wave 1. At Wave 2 EF was assessed using conflict inhibition task (Arrows task, Davidson et al., 2006), a shifting task (Smiling Face task, Huizinga et al., 2006) and WM task (Thorndike et al., 1986). Concurrent correlations between ToM and EF were of moderate strength at age 6 but weak at age 10. Cross-lagged longitudinal analyses revealed that developmental relations between these two constructs were non-significant ($\beta s < .15$). Devine et al. (2016) proposed that the relations between ToM and EF may change in nature beyond the preschool years. These results may indicate that EF predicts ToM development in preschool but not in middle childhood. However, two methodological features of this study, namely the extended interval between time points and use of different measures at each wave of the study, suggest an alternative interpretation. The authors acknowledged that as well as attenuating any correlations between the two constructs over time, the extended temporal span and the use of different measures might have masked any proximal developmental influence of EF on ToM. By adopting a short-term longitudinal design in which the same tasks were used at each time-point, we hoped to identify whether Devine et al.'s (2016) findings reflect a genuine developmental shift in links between EF and ToM or whether this contrast is a spurious artefact of methodological factors.

In sum, data from the only two longitudinal studies of the links between EF and ToM in middle childhood have produced equivocal results. Our third aim was therefore to conduct cross-lagged analyses of the associations between two aspects of executive function (i.e., IC and WM) and performance on a battery of reliable, valid ToM tasks that has previously proved sensitive to individual differences (Devine & Hughes, 2016). In contrast with previous studies, we focused on children within a relatively narrow age-range, monitored

performance at 6-month intervals and used the same measures of EF and ToM at each time point.

Summary of Aims

The current study had three distinct aims. The first aim was to examine, for the first time, the *development and longitudinal rank-order stability* of ToM, WM and IC across three time points over a 12-month period in middle childhood. The second aim was to examine the *concurrent relations* between ToM, IC and WM in middle childhood. The third and final aim of our study was to assess the *developmental relations* between EF and ToM. We designed a three-wave cross-lagged longitudinal study in which we measured individual differences in children's ToM and EF across three time points from age 9.5 to 10.5 years. This design enabled us to examine the direction of the relations between EF and ToM, after controlling for rank-order stability and language ability.

Method

Participants

The participants ($N = 113$; 61 boys) were recruited from 5 public schools (9 classes) in Northern Italy. The children were 9.36 years old ($M = 112.3$ months, $SD = 4.18$ months) at Time 1, 9.85 years old ($M = 118.3$ months, $SD = 4.15$ months) at Time 2, and 10.39 years old ($M = 124.7$ months, $SD = 4.06$ months) at Time 3. The measurement intervals from Time 1 to Time 2, $M = 6$ months, $SD = 0.72$ months, and from Time 2 to Time 3, $M = 6.26$, $SD = 0.74$, differed significantly, $M_{\text{Diff}} = 0.25$ months, $SE = 0.12$, $t(112) = 2.08$, $p = .04$, $d = 0.20$. All the children were fluent Italian speakers and none had a history of developmental delay or learning disorder. With regard to family structure: 21 subjects were singletons, 63 had one sibling, 22 had two siblings, 5 had three siblings, and 2 had more than 3 siblings.

Measures

Socio-economic status. We used the Family Affluence Scale (FAS, Currie et al., 2008) to assess material wealth. The FAS is a child-report questionnaire consisting of four questions evaluating family car ownership (range 0-2), having an unshared bedroom (range 0-1), the number of computers at home (range 0-3), and the number of family vacations in a 12 month period (range 0-3). Responses were summed into an overall index of family socio-economic background. Possible total scores could range from 0 to 9 points.

Verbal ability. We administered the Italian version of the vocabulary subtest of the Primary Mental Abilities test at each time point in this study (PMA; Thurstone & Thurstone, 1962; Rubini & Rossi, 1982). This test requires children to select the right synonym of a target word out of 4 alternatives. There were 30 target words and a time limit of 7 min was set. Possible total scores could range from 0 to 30 points.

WM. We used the Backwards Digit Span task (Orsini, 1997). This task required children to listen to 7 sequences of 2 to 8 digits and recall them in reverse order. Each sequence was marked as a pass (1) or fail (0), and total scores could range from 0 to 7 points.

IC. We administered the Arrows task (Davidson et al., 2006). For each trial, a single arrow was presented on either the left or the right of a computer screen. On congruent trials, the arrow pointed straight down (towards the response button on the same side as the arrow); on incongruent trials, the arrow pointed diagonally towards the opposite side at a 45° angle (toward the response button on the side opposite to the arrow). The incongruent condition required inhibiting the tendency to respond on the same side where the stimulus occurred. Stimulus presentation time was 750ms and the inter-stimulus interval was 500ms. Congruent and Incongruent trials were presented in 6 blocks of 4 stimuli each (congruent right, congruent left, incongruent right and incongruent left). Each block presented the stimuli in randomized order with no breaks between the blocks. We computed an accuracy score for

the incongruent trials (Devine et al., 2016) by summing the total number of correct responses. Anticipatory responses ($< 200ms$) and missed responses were scored as errors.

ToM. We administered the Strange Stories Test (Happé, 1994; White et al., 2009) and the Silent Films task (Devine & Hughes, 2013). These two tasks have been used in large-scale studies of individual differences in children's ToM across middle childhood (Devine & Hughes, 2013; 2016). They show excellent one-month test-retest reliability in middle childhood, high levels of internal consistency (Devine & Hughes, 2016) and convergent validity (Devine & Hughes, 2016). Moreover, performance on both tasks shows construct validity in that individual differences in performance on both tasks at age 10 are significantly associated with individual differences in children's false belief understanding at age 6 (Devine et al., 2016), children's self-rated social exclusion (Devine & Hughes, 2013) and teachers' ratings of children's social competence (Devine et al., 2016).

The Strange Stories requires children to interpret non-literal statements by making inferences about mental states (see Appendix A). We administered five stories: 2 double bluff, 1 persuasion, 1 misunderstanding, 1 white lie. They were presented in a written form, the experimenter read each of them and asked children to write down their answers (while having the stories in front of them) and no time limit was set. In line with the scoring procedures (White et al., 2009), we rated children's responses using a three-point scale: 0 for an incorrect or did not know answer, 1 for partial or implicit correct answers and 2 for full and explicit correct answers. Total scores could range from 0 to 10 points. A second rater independently coded 25% of the responses at each time point and inter-rater agreement was established using Cohen's kappa (at Time 1, $\kappa = .93$; at Time 2, $\kappa = .95$; at Time 3, $\kappa = .77$).

The Silent Film task measures individual differences in participants' use of their understanding of desires, knowledge and beliefs to explain social situations (see Appendix B). The Silent Film task consists of 5 short film clips from a classic silent comedy, Harold

Lloyd's 'Safety Last' (Roach, Newmeyer, & Taylor, 1923). The children were instructed to watch each short film clip once and to answer a question about a character's behavior. Oral answers were transcribed and scored using a three-point scale according to the standard scoring guidelines: 0 for an incorrect response, 1 for a partially correct response and 2 for an explicit mentalistic explanation of the character's behavior. There were 5 film clips and 6 questions. Total scores could range from 0 to 12 points. A second rater independently coded 25% of the responses at each time point and inter-rater agreement was established using Cohen's kappa (at Time 1, $\kappa = .91$; at Time 2, $\kappa = .92$; at Time 3, $\kappa = .85$).

Procedure

Children's ToM, EF, and verbal ability were measured at three time points: at the beginning of Year 4 (Time 1), at the end of Year 4 (Time 2) and, finally at the beginning of Year 5 (Time 3)¹. Whole class sessions were used to administer the test of verbal ability and the strange stories. Children were cautioned not to discuss the items and to work independently; no feedbacks were provided. All the other tasks were administered individually in a quiet room located in the school. We also gathered information about children's socio-economic background and number of siblings. Parental written informed consent was obtained. The order of tasks was counterbalanced.

Results

Data Reduction and Descriptive Statistics

We used structural equation modelling in *MPlus* Version 7 (Muthèn & Muthèn, 2012) to examine the longitudinal associations between measures of EF and individual differences in ToM. Table 1 and Table 2 show the descriptive statistics and correlations between measures of ToM and EF across the three time points of the study.

¹ Please note that school years refer to the Italian school system in which children start school at about 6 years of age

Table 1

Descriptive Statistics

	Time 1					Time 2					Time 3				
	VA	SF	SS	IC	WM	VA	SF	SS	IC	WM	VA	SF	SS	IC	WM
<i>M</i>	26.31	7.36	5.12	2.45	2.67	27.56	8.26	6.82	2.67	2.61	27.99	8.81	7.26	4.73	2.72
<i>SD</i>	3.64	2.01	1.69	2.19	1.09	2.92	1.73	1.82	2.19	0.87	2.93	1.74	1.58	3.12	0.95
Min	12	2	1	0	1	11	4	1	0	1	9	4	3	0	1
Max	30	11	10	9	5	30	12	10	9	5	30	12	10	12	7

Note. VA = Verbal Ability; SF = Silent Film Task; SS = Strange Stories Task; IC = Inhibition; WM = Working Memory.

Table 2

Correlations and Partial Correlations (Above Diagonal) Between Study Measures

	1	2	3	4	5	6	7	8	9	10	11	12
Age (Concurrent)	.04	-.05	.12	-.01	-.05	-.01	.10	.04	-.01	-.06	-.02	-.01
Gender	-.01	.02	-.03	.02	.04	-.21 ⁺	.03	.03	.10	.09	.11	-.05
FAS	.05	.12	.02	-.08	.13	.05	.03	.03	-.01	.03	-.02	.20 ⁺
Siblings	.13	.02	.02	-.21 ⁺	.09	.08	.01	-.01	.04	-.06	-.05	-.22 ⁺
Verbal Ability (Concurrent)	.14	.39 [*]	.21	.16	.29 [*]	.39 [*]	.25 [*]	.17	.16	.48 [*]	.26 [*]	.12
1. Silent Film T1	-	.21 ⁺	.06	.03	.43 [*]	.12	.16	.05	.29 [*]	.20 ⁺	.11	.14
2. Strange Stories T1	.24 [*]	-	-.05	.05	.13	.40 [*]	.08	.15	.21 ⁺	.39 [*]	.07	.03
3. Working Memory T1	.09	.03	-	.14	.13	.13	.49 [*]	.23 ⁺	.16	.22 ⁺	.47 [*]	.18
4. Inhibition T1	.05	.11	.16	-	-.22 ⁺	-.01	.12	.37 [*]	-.10	.01	-.05	.34 [*]
5. Silent Film T2	.45 [*]	.22 ⁺	.17	-.17	-	.21 ⁺	.04	-.04	.40 [*]	.14	.20 ⁺	.08
6. Strange Stories T2	.16	.49 [*]	.20 ⁺	.05	.30 [*]	-	.08	-.01	.22 ⁺	.33 [*]	-.01	.04
7. Working Memory T2	.19 ⁺	.15	.52 [*]	.14	.11	.17	-	.20 ⁺	.14	.32 [*]	.44 [*]	.14
8. Inhibition T2	.08	.22 ⁺	.27 [*]	.39 [*]	.01	.06	.24 [*]	-	-.03	-.04	.01	.50 [*]
9. Silent Film T3	.31 [*]	.26 [*]	.19 ⁺	-.07	.44 [*]	.29 [*]	.19 ⁺	.01	-	.21 ⁺	.16	.12
10. Strange Stories T3	.23 [*]	.48 [*]	.27 [*]	.07	.25 [*]	.45 [*]	.39 [*]	.04	.27 [*]	-	.15	.13
11. Working Memory T3	.13	.16	.49 [*]	-.01	.26 [*]	.09	.47 [*]	.06	.19 ⁺	.25 [*]	-	.01
12. Inhibition T3	.16	.10	.21 ⁺	.35 [*]	.13	.11	.18	.52 [*]	.14	.17	.04	-

Note. * $p < .01$. ⁺ $p < .05$. T1 = Time 1. T2 = Time 2. T3 = Time 3. Correlations with age and verbal ability use concurrent age and verbal ability scores. Partial correlations control for concurrent verbal ability and age. Longitudinal partial correlations control for initial verbal ability and age at the last time point.

On the basis of preliminary correlational and confirmatory factor analysis (see Appendix C), our primary analyses were conducted using an aggregate ToM measure. The measures of IC and WM were not correlated at Time 1 or Time 3 and only weakly correlated at Time 2. We therefore examined the relations between ToM and these two variables separately.

Development of Theory of Mind and Executive Function Across Middle Childhood

First, we examined change in performance on each of the key study variables. Performance on the Strange Stories, $F(2,224) = 93.81, p < .001$, partial $\eta^2 = .46$, and the Silent Film task, $F(2,224) = 29.33, p < .001$, partial $\eta^2 = .21$, improved significantly across the three waves of the study. On the Strange Stories there were significant gains from Time 1 to Time 2, $t(112) = 10.16, p < .001, d = 0.96$, from Time 2 to Time 3, $t(112) = 2.57, p < .05, d = 0.24$, and from Time 1 to Time 3, $t(112) = 10.16, p < .001, d = 0.96$. There were also significant gains from Time 1 to Time 2, $t(112) = 4.79, p < .001, d = 0.89$, and from Time 2 to Time 3, $t(112) = 3.18, p < .01, d = 0.30$, and from Time 1 to Time 3, $t(112) = 6.90, p < .001, d = 0.65$, on the Silent Film task. Inspection of the item-level data (Table S1²) revealed that performance on all items improved significantly from Time 1 to Time 3 suggesting that mean changes in performance were not driven by any one item in particular. Unsurprisingly, scores on the aggregate ToM variable scores also improved significantly across the three waves of the study, $F(2, 224) = 101.72, p < .001, partial \eta^2 = .48$. There were significant gains in performance from Time 1 to Time 2, $t(112) = 9.66, p < .001, d = 0.91$, from Time 1 to Time 3, $t(112) = 13.56, p < .001, d = 1.27$, and from Time 2 to Time 3, $t(112) = 4.02, p < .001, d = 0.38$. Mauchly's test revealed that the assumption of sphericity had been violated for the effects of time on IC task and so we corrected the degrees of freedom using the Greenhouse-Geisser correction. There was a significant main effect of time on performance

² Supplementary material

across the three waves of the study, $F(1.85, 206.72) = 46.76, p < .001, \text{partial } \eta^2 = .30$. Post-hoc tests revealed that accuracy on the IC task was significantly greater at Time 3 than at Time 2, $t(112) = 8.04, p < .001, d = 0.76$, and Time 1, $t(112) = 7.80, p < .001, d = 0.73$. There were no differences in performance between Time 1 and Time 2, $t(112) = 0.97, p = .33, d = 0.09$. There was no evidence of time-related change in performance on the WM task, $F(2, 224) = 0.66, p = .52, \text{partial } \eta^2 = .01$.

Concurrent Relations Between Theory of Mind and Executive Function

Table 2 shows the correlations and partial correlations (controlling for verbal ability and age) between each of the key study measures and background variables. Visual inspection of the correlation matrix showed that there were no differences in the pattern of concurrent associations between each individual measure of ToM and IC or WM. Thus, here we describe the correlations between the aggregate measure of ToM and measures of EF within each time point.

Within each time point, individual differences in performance on the aggregate measure of ToM were significantly correlated with concurrent measures of verbal ability, $r_{\text{Time 1}}(111) = .32, p < .01, r_{\text{Time 2}}(111) = .43, p < .01, \text{ and } r_{\text{Time 3}}(111) = .40, p < .01$. There were no significant correlations between ToM and gender, $-.11 \leq r(111) \leq .12$, family affluence, $.01 \leq r(111) \leq .11$, or number of siblings, $-.01 \leq r(111) \leq .10$. Even when the effects of earlier verbal ability and later age were taken into account, there was moderate rank-order stability in performance on the aggregate measure of ToM from Time 1 to Time 2, $r(109) = .45, p < .01$, from Time 1 to Time 3, $r(109) = .44, p < .01$, and from Time 2 to Time 3, $r(109) = .46, p < .01$.

At Time 1 and Time 2 there were no significant concurrent associations between individual differences in performance on the aggregate measure of ToM and either WM, $r_{\text{Time 1}}(111) = .08, r_{\text{Time 2}}(111) = .17$, or IC, $r_{\text{Time 1}}(111) = .10, r_{\text{Time 2}}(111) = .04$. In contrast,

there were weak associations between individual differences in ToM and both WM, $r(111) = .28, p < .01$, and IC, $r(111) = .19, p < .05$, at Time 3. The Time 3 correlation between WM and ToM remained significant even when the effects of concurrent verbal ability and age were taken into account, $pr(109) = .20, p < .05$. The Time 3 correlation between IC and ToM was weak and non-significant, $pr(109) = .16, p < .10$.

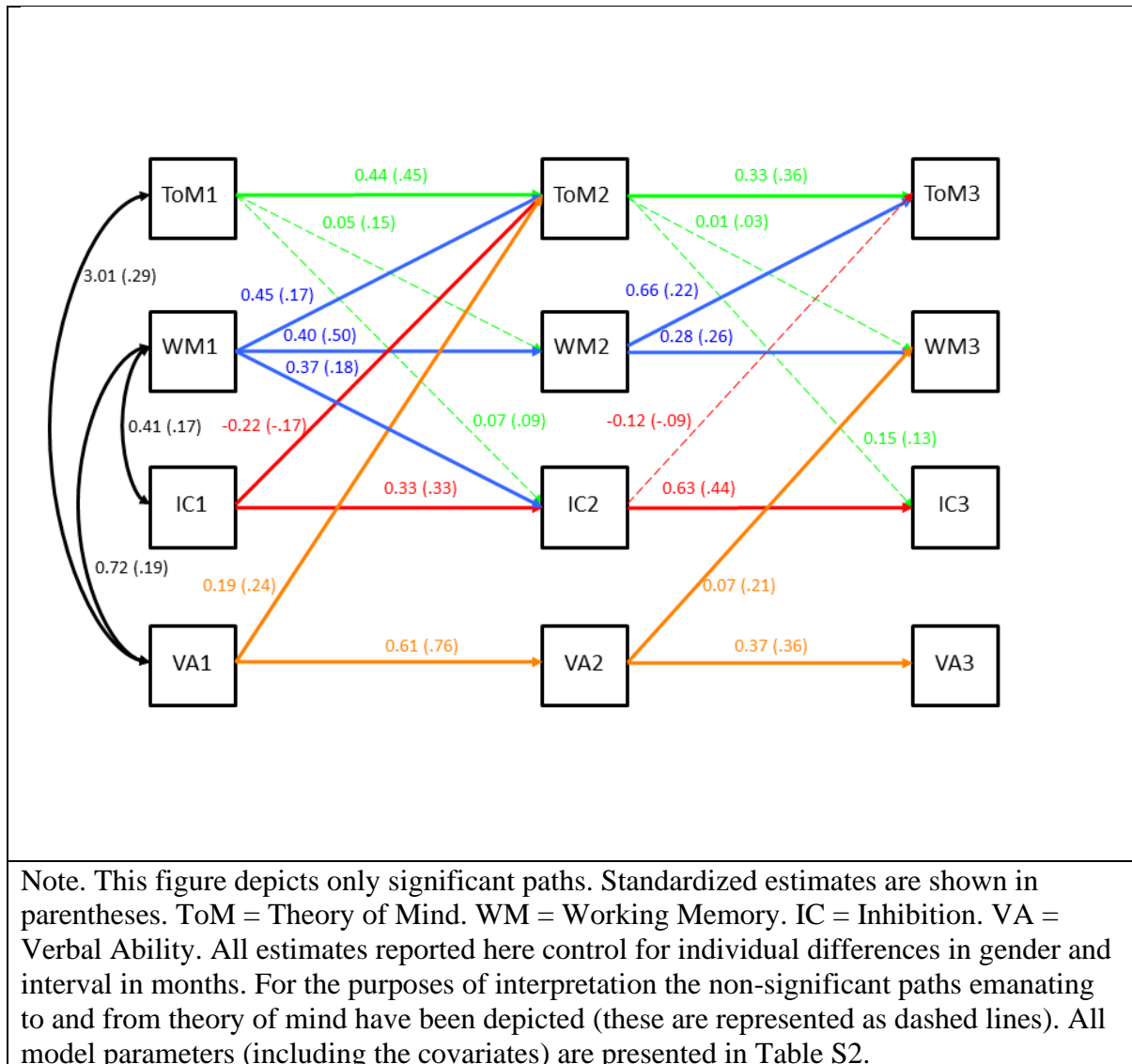
Longitudinal Relations Between Theory of Mind and Executive Function

We used four primary criteria to assess the acceptability of our structural models: a non-significant χ^2 test of model fit, Comparative Fit Index (CFI) > 0.95 , Tucker Lewis Index (TLI) > 0.95 and Root Mean Square Error of Approximation (RMSEA) < 0.08 (Brown, 2015). In each of our longitudinal models we estimated the stability of each construct over time by regressing Time 3 measures onto Time 2 measures and Time 2 measures onto Time 1 measures. We regressed the Time 3 and Time 2 measures onto individual differences in prior verbal ability and gender to control for these variables. We also regressed the Time 2 values onto the Time 1 to Time 2 interval (in months) and the Time 3 values onto the Time 2 to Time 3 interval to control for individual differences in the length of time between each testing period. Crucially, we estimated cross-lagged paths to assess the direction of the developmental relations between our measures of executive function and ToM. Finally, we permitted each of the predictor variables (and the outcome variables) in our three-wave model to co-vary. Given the non-normal distribution of some of the variables in our dataset, we used robust maximum likelihood (MLM) estimation with a mean-adjusted χ^2 test statistic for each of our models (Kline, 2011).

Our cross-lagged longitudinal model provided an acceptable fit to the data, $\chi^2(31) = 29.50, p = 0.54$, CFI = 1.00, TLI = 1.01, RMSEA = 0, 90% CI [0, 0.07]. A simplified depiction of this model (showing only significant pathways) is shown in Figure 1. For the

purposes of interpretation, the non-significant paths emanating to and from ToM have been included as dashed lines.

Figure 1. Simplified Cross-Lagged Longitudinal Model.



The complete results for this model are presented in Table S2³. At Time 2 the model accounted for 40% of the variance in ToM, 22% of the variance in IC and 30% of the

³ Supplementary material

variance in WM. At Time 3 the model explained 43% of the variance in ToM, 28% of the variance in IC and 23% of the variance in WM.

Four features of this model deserve note. First, all four variables measured (i.e., ToM, WM, IC and verbal ability) exhibited moderate rank-order stability over time. Second, there were significant cross-lagged associations between Time 1 WM and Time 2 ToM and between Time 2 WM and Time 3 ToM (depicted in blue in Figure 1). In contrast there were no significant cross-lagged relations between earlier ToM and later WM (depicted as green dashed lines in Figure 1). Third, with the exception of the path between Time 1 IC and Time 2 ToM (depicted in red in Figure 1), there were no other cross-lagged associations between IC and ToM (depicted as red dashed lines in Figure 1). Fourth, with the exception of the path between Time 1 WM and Time 2 IC, there were no significant cross-lagged relations between WM and IC. We used model comparisons to investigate these results further (see Appendix D).

Discussion

The main aims of the present 12-month longitudinal study were to examine the *development* of ToM and EF in middle childhood and assess both concurrent and across-time relations between individual differences in ToM and EF. Our study yielded three key sets of results. First, performance on each measure of ToM (and the aggregate theory-of-mind score) was stable and improved over the course of the study. Supplementary analyses showed that these changes are unlikely to be due to any one item in either test of ToM. With regard to the measures of EF, children's IC performance, but not their WM, improved across the course of the study. Second, correlation analyses showed that there were no concurrent associations between measures of ToM, IC or WM at Time 1 and Time 2 and weak associations between these variables at Time 3. Third, our longitudinal structural equation models (controlling for verbal ability, gender and individual differences in the interval

duration) demonstrated that there were unique longitudinal associations between WM and later ToM and that between Time 2 and Time 3 this association was unidirectional. In contrast, there was only a weak negative association between Time 1 IC and Time 2 ToM but not vice versa.

Theory of Mind in Middle Childhood

The present study follows a growing interest in ToM (Hughes, 2016). Our longitudinal study contributes to the emerging literature on ToM in middle childhood in at least three ways. First, our study revealed a steady increase in children's performance on the Strange Stories (White et al., 2009) and Silent Films (Devine & Hughes, 2013) tasks in late middle childhood. This result supports the hypothesis that ToM development in middle childhood is characterized not by any stage-like conceptual change, but rather by a gradual improvement in children's ability to apply their ToM skills to a variety of complex social situations (Apperly, 2011).

Second, in line with existing findings (Banerjee et al., 2011; Lecce et al., 2010; Devine et al., 2016) our results provide further evidence of strong stability of individual differences in ToM in middle childhood. This stability is important for more than one reason. Methodologically, it reassures us of the reliability of our chosen measures. Theoretically, it demonstrates that individual differences are not temporary fluctuations in task performance but, rather, genuine characteristics of the children (see Hughes & Devine, 2015b). These results add to the growing body of evidence suggesting that individual differences in ToM are meaningful for children's social (Banerjee et al., 2011; Devine & Hughes, 2013; Devine et al., 2016) and cognitive development (Lecce, Caputi, et al., 2014) in primary school years. Here it is worth noting that ToM is an ability embedded in real-life school activities. It underpins children's awareness of what is happening and potentially influences children's behavior during social exchanges and cognitive activities in classes and play time.

Executive Function in Middle Childhood

Results on EF showed, between the ages 10 (Time 2) and 10.5 years (Time 3), a substantial increase in children's IC but not in their WM. No change in IC between Time 1 and Time 2 was found. The results for IC add to a growing body of literature (Brocki & Bohlin, 2004; Cragg & Nation, 2008; Huizinga et al., 2006; Jonkman, 2006; Lamm, Zelazo, & Lewis, 2006). The absence of age-related differences in WM does not mean that no changes in WM occur during middle childhood as shown by the literature (Brocki & Bohlin; 2004; Prencipe et al., 2011). It may indicate that the development of WM is discontinuous and the age range we considered in the present study (9.5 -10.5 years) is not very revealing for developmental shifts. We found no evidence for an association between WM and IC. This may reflect the specific characteristics of the measures that we selected. Davidson et al. (2006) designed the Arrows IC task to have relatively few WM demands. Indeed, the Arrows Task requires inhibiting the tendency to respond on the same side as the stimulus when a diagonal arrow appears, but it requires little or no WM, as the arrow points directly to the correct response button on all trials. The lack of association between IC and WM may also reflect the increased diversity in the structure of EF associated with growing age as reported by Shing and colleagues (2010) and Huizinga et al. (2006) in children aged 4 to 14,5 and 7 to 21 years respectively.

Relations Between ToM and EF in Middle Childhood

Whereas in preschoolers a moderate cross-sectional correlation between EF and ToM and a developmental association between early EF and later ToM task performance (Devine & Hughes, 2014) were found, studies on middle childhood have provided equivocal results (Austin et al., 2014; Devine et al., 2016). To address the methodological factors that may have contributed to these equivocal findings, the present study used the same ToM and EF tasks at each time point and examined longitudinal associations between ToM and EF across

short temporal periods (6-months) in a narrowly age defined group of children. Consistent with data from the preschool literature (Devine & Hughes, 2014), we found a uni-directional developmental relation between ToM and one aspect of EF, WM: early individual differences in WM predicted later individual differences in ToM. However, there were overall no developmental relations between ToM and IC across the three time points covered in our sample. In other words WM, but not IC, appears to contribute to the on-going development of ToM in late middle childhood.

Interestingly, in our study WM was related to later, but not concurrent, ToM. This finding corroborates results of a number of studies involving school-aged children and using the same measures as in the present study: Strange Stories (Bock, Gallaway, & Hund, 2015; Osterhaus, Koerber, & Sodian, 2016; Oswald, 2013) and Silent films (Wang et al., 2016). Together these null findings contrast with studies of pre-schoolers that indicate moderate correlations between measures of WM and false belief understanding (e.g., Devine & Hughes, 2014). This contrast is open to at least two explanations.

First, traditional false-belief tasks differ from these ‘advanced’ measures of ToM in obvious ways. Specifically, the classic change-of-location false belief task requires children to keep track of multiple story elements. In contrast, in the Strange Stories task the text remains available to the child and so minimizes WM demands. Likewise, the Silent Film task has minimal WM demands because the clips are very brief (approximately 30s long) and are non-verbal. Each of the scenarios in this task contains a comparable number of story elements to the classic change-of-location false belief task. From a developmental perspective, while this number of story elements could conceivably place demands on WM in a pre-schooler, it is unlikely to challenge the WM capacity of school-aged child.

Second, an imaging study conducted by Blakemore and colleagues (Blakemore, den Ouden, Choudhury, & Frith, 2007) has shown that when tested using a vignette-based

measure of belief reasoning, adolescents and adults show comparable behavioural performance but this performance is underpinned by different neural mechanisms. Specifically, relative to adults, adolescents exhibited significantly greater activation of the medial prefrontal cortex and lower levels of activation in the right superior temporal sulcus. While preliminary, these findings underscore the need for greater attention to developmental processes when studying the relations between ToM and EF. Results of the present study also fit with those of a recent training study on children of the same age range (Lecce & Bianco, 2017). In this study children with a better WM at pre-test were more able to capitalize on a ToM conversational-based training than children with low WM. Despite we cannot be sure that learning ToM through training leverages the same cognitive processes that promote ToM in daily life, it is important to note that the training activities used in Lecce and Bianco training (i.e., conversations on narratives) were selected to be very similar to those children encounter at school during daily lectures.

The longitudinal relationship between WM at Time 1 and Time 2 with ToM at Time 2 and Time 3 (respectively) is in line with the *emergence* account. Whereas this account is tolerant to a concurrent dissociation between EF and ToM, it posits longitudinal relations between early EF and later ToM because EF is needed for ToM to develop. Although the present study did not allowed us to address the issue of *why* this longitudinal association emerged, we speculate the WM might facilitate children in the process of learning about other minds from relevant social experience. Specifically, WM is likely to promote the ability to notice discrepancies between previously established expectations/knowledge and subsequent inputs and, thus, to flexibly update prior knowledge and mental models (Benson & Sabbagh, 2013). This, in turn, is likely to promote the development of ToM over time. Crucially, the role of WM seems particularly important in middle childhood as social

situations are complex in this developmental period and children need to hold in mind and keep track of multiple and often discordant perspectives (Oswald, 2013).

Our findings go some way towards clarifying the findings reported by Devine et al. (2016). Their study spanned over 4 years in middle childhood and revealed no developmental links between early EF (measured at age 6) and later ToM (measured at age 10). Devine et al. (2016) suggested that the nature of the relations between EF and ToM shifts with development. That is, EF may contribute to the emergence of ToM in preschool but to the expression of ToM in middle childhood. While our findings do not directly support this hypothesis, they do suggest a developmental shift in the nature of the relations between ToM and EF. In particular, whereas there appears to be a general link between a range of different measures of EF in the preschool years (Devine & Hughes, 2014), our results suggest that in late middle childhood the links with ToM might be specific to WM.

Our findings show that the role of IC in ToM development in middle childhood is minor. We indeed, found only one significant link out of 6; this was a significant negative effect of T1 IC on T2 ToM. Even if we are aware that this data requires further replication, the unexpected direction of this association deserves some attention. A large amount of studies conducted on preschoolers indeed found a positive (instead of negative) longitudinal association between early IC and later ToM (Devine & Hughes, 2014). This relationship was due to the fact that in order to pass first-order FB scenarios, the child is required to inhibit a predominant response to answer based on her own accurate perspective of reality. It is possible that this mechanism does not work for more advanced ToM understanding such as the one considered in the present study. Indeed, children's knowledge may be helpful in order to respond to some Strange Stories and Silent Films. It is possible that children rely, at least in part, on their previous social knowledge in order to perform these tasks. For example, to understand a white lie or a persuasion scenario children's previous experience

may have created a structured representation that helps them in decoding social information.

Following this line of interpretation, higher IC may hinder, rather than promote, the understanding of complex social scenarios involving mental states.

Caveats and Conclusions

Overall our study showed that although WM, IC and ToM are not concurrently correlated in middle childhood, individual differences in WM contribute to the development, at least in the short term of ToM. Our study has a number of limitations that should be acknowledged and that correspond with directions for further research. First, we used only one task to index IC and WM; aggregate measures of IC and WM would provide more reliable estimates of EF in future research. Second larger samples are needed to confirm our results and study how change in one variable drives change in the other. That said, our study is the first to examine the development and relations between EF and ToM in middle childhood using a short-term multi-wave longitudinal design and has several strengths. First, the independence of our measures of WM and IC meant that we were able to assess their unique contributions to individual differences in ToM in middle childhood and reveal specific pattern of associations across time. Second, the short-term longitudinal design of the study and the relatively narrow age range of the participants minimized the effect of potential confounding factors, enabling us to draw a specific picture of the relations between ToM and EF in middle childhood. We were able to show that WM plays a significant role in supporting the child to reach a more mature understanding of mental phenomena in a 6 months span, but more research is needed to unfold the specific mechanisms explaining this effect. Indeed, our design did not permit us to study whether this developmental relation is specific to ToM development. Similarly, future studies should manipulate the executive demands of the focal task to check if different EF levels have an impact on the EF-ToM

developmental relation. This manipulation would also clarify if, during middle childhood, EF has also a role in ToM expression.

Future research should also address the issue of the social mechanisms underlying the associations between WM and ToM (Benson & Sabbagh, 2013). One interesting possibility is that WM might enable children to learn from the types of everyday experiences that provide them with information about other people's minds (Benson & Sabbagh, 2013). More specifically, Benson and Sabbagh (2009) suggest that WM may have a double role in enabling children to: (i) elicit and maintain social interactions from which they pick up new insights on mental states; and (ii) make use of ToM-relevant information.

Our study extends existing work on the relations between ToM and EF in middle childhood and has both practical and theoretical implications. In view of the fact that EF has a significant, but rather weak effect on later ToM, our study may provide an impetus for investigating the role of social factors, such as peer relationships (Banerjee et al., 2011) and conversations (Lecce, Bianco, et al., 2014), in ToM development in middle childhood. Indeed, social environmental factors in general (e.g., friends and peers) account for 66% (Ronald, Viding, Happé, & Plomin, 2006), and, more precisely, conversations about the mind for about 23%, of variance in ToM performance during middle childhood (Lecce et al., 2014). More integrative work that examine the relative salience and interplay between the predictors of both EF and ToM would be a fruitful avenue for future research in this area.

References

- Ahmed, F. S., & Miller, S. L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism & Developmental Disorders, 41*, 667-678. doi:10.1007/s10803-010-1087-7
- Amsterlaw, J., Lagattuta, K. H., & Meltzoff, A. N. (2009). Young children's reasoning about the effects of emotional and physiological states on academic performance. *Child Development, 80*, 115–133. doi:10.1111/j.1467-8624.2008.01249.x
- Apperly, I. A. (2011). *Mindreaders: The cognitive basis of 'theory of mind'*. New York, NY, USA: Psychology Press.
- Archibald, S. J. , & Kerns, K. A.(1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychol, 5*, 115–129. doi:10.1076/chin.5.2.115.3167
- Austin, G., Groppe, K., & Elsner, B. (2014). The reciprocal relationship between executive function and theory of mind in middle childhood: A 1-year longitudinal perspective. *Frontiers in Psychology, 5*, 1-11. doi:10.3389/fpsyg.2014.00655
- Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development, 82*, 1887-1905. doi:10.1111/j.1467-8624.2011.01669.x
- Baron-Cohen, S., O' Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 29*, 407-418. doi:10.1023/A:1023035012436
- Bender, P. K., Pons, F., Harris, P. L., & de Rosnay, M. (2011). Do young children misunderstand their own emotions?. *European Journal of Developmental Psychology, 8*, 331–348. doi:10.1080/17405629.2010.495615

- Benson, J. E., & Sabbagh, M. A. (2010). Theory of mind and executive functioning: A developmental neuropsychological approach. In , P. D. Zelazo, M. Chandler, E. Crone (Eds.) , *Developmental social cognitive neuroscience* (pp. 63-80). New York, NY, US: Psychology Press.
- Benson, J., & Sabbagh, M. A. (2013). The relation between executive functioning and social cognition. *Executive functions*, 22, 1-5. Retrieved from <http://www.child-encyclopedia.com/executive-functions/according-experts/relation-between-executive-functioning-and-social-cognition>
- Bianco, F., Lecce, S., & Banerjee, R. (2016). Conversations about mental states and theory of mind development in middle childhood: A training study. *Journal of Experimental Child Psychology*, 149, 41-61. doi:10.1016/j.jecp.2015.11.006
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78, 647-663. doi:10.1111/j.1467-8624.2007.01019.x
- Blakemore, S., den Ouden, H., Choudhury, S., & Frith, C. (2007). Adolescent development of the neural circuitry for thinking about intentions. *Social Cognitive and Affective Neuroscience*, 2, 130-139. doi:10.1093/scan/nsm009
- Bock, A. M., Gallaway, K. C., & Hund, A. M. (2015). Specifying links between executive functioning and theory of mind during middle childhood: Cognitive flexibility predicts social understanding. *Journal of Cognition and Development*, 16, 509-521. doi:10.1080/15248372.2014.888350
- Brocki, K. G., & Bohlin G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, 26, 571-593. doi:10.1207/s15326942dn2602_3

Brown, T. A. (2015). *Confirmatory factor analysis for applied research., 2nd ed.* New York, NY, US: Guilford Press.

Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson-McInnis, M. S., & Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *Journal of Experimental Child Psychology, 146*, 66-78. doi:10.1016/j.jecp.2016.01.014

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032-1053. doi:10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant & Child Development, 11*, 73-92. doi:10.1002/icd.298

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12*, 314-325. doi:10.1006/nimg.2000.0612

Cragg, L. K., & Nation, K. (2008). Go or no-go? Developmental improvements in the efficiency of response inhibition in mid-childhood. *Developmental Science, 11*, 819-827. doi:10.1111/j.1467-7687.2008.00730.x

Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behavior in School-aged Children (HBSC) family affluence scale. *Social Science & Medicine, 66*, 1429-36. doi:10.1016/j.socscimed.2007.11.024

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from

manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037-2078. doi:10.1016/j.neuropsychologia.2006.02.006

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of mind, gender and social experiences in middle childhood. *Child Development*, *84*, 989-1003. doi:10.1111/cdev.12017

Devine, R. T., & Hughes C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*, 1777-1794. doi:10.1111/cdev.12237

Devine, R.T., & Hughes, C. (2016). Measuring theory of mind in middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, *149*, 23-40. doi:10.1016/j.jecp.2015.07.011

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, *52*, 758-771. doi:10.1037/dev0000105

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168. doi:10.1146/annurev-psych-113011-143750

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2–7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153. doi:10.1016/0010-0277(94)90068-X

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129-154. doi:10.1007/BF02172093

- Henning, A., Hofer, T., & Aschersleben, G. (2012). Erweiterte "Theory of Mind- Skala" für 3- bis 11-jährige Kinder. In *Extended Theory-of-Mind Scale*, trans. Peterson, Wellman, & Slaughter. Saarbrücken: Universität des Saarlandes.
- Henning, A., Spinath, F. M., & Aschersleben, G. (2011). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology*, *108*, 513-531. doi:10.1016/j.jecp.2010.10.006
- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*, 233-253.
doi:10.1111/j.2044-835X.1998.tb00921.x
- Hughes, C. (Ed). (2016). Theory of mind grows up: Reflection on new research on theory of mind in middle childhood and adolescence [Special issue]. *Journal of experimental Child Psychology*, *149*, 1-5. doi:10.1016/j.jecp.2016.01.017
- Hughes, C., & Devine, R. T. (2015a). Individual differences in theory of mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives*, *9*, 149-153. doi:10.1111/cdep.12124
- Hughes, C. & Devine, R.T. (2015b). Family influences on theory of mind: A review. To appear in V. Slaughter and M. De Rosnay (Eds.), *Environmental Influences on Theory of Mind*. Hove, UK: Psychology Press.
- Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology*, *28*, 645-668.
doi:10.1207/s15326942dn2802_5
- Hughes, C., Ensor, R., & Wilson, A., (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, *35*, 20-36. doi:10.1080/87565640903325691

- Huizinga, M., Dolan, C. V., & Van Der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*, 2017–2036. doi:10.1016/j.neuropsychologia.2006.01
- Im-Bolter, N., Agostino, A., & Owens-Jaffray, K. (2016). Theory of mind in middle childhood and early adolescence: Different from before? *Journal of Experimental Child Psychology*, *149*, 98-115. doi:10.1016/j.jecp.2015.12.006
- Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology*, *52*, 1192-1205. doi:10.1037/dev0000140
- Jacobson, L. A., Williford, A. P., & Pianta, R. C. (2011). The role of executive function in children's competent adjustment to middle school. *Child Neuropsychology*, *17*, 255-280. doi:10.1080/09297049.2010.535654
- Jonkman, L. (2006). The development of preparation, conflict monitoring and inhibition from early childhood to young adulthood; A Go/Nogo ERP study. *Brain Research*, *1097*, 181-193. doi:10.1016/j.brainres.2006.04.064
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. New York, USA: The Guilford Press.
- Lagattuta, K. H. (2014). Linking past, present, and future: Children's ability to connect mental states and emotions across time. *Child Development Perspectives*, *8*, 90–95. doi:10.1111/cdep.12065
- Lagattuta, K. H., Elrod, N. M., & Kramer, H. J. (2016). How do thoughts, emotions, and decisions align? A new way to examine theory of mind during middle childhood and beyond. *Journal of Experimental Child Psychology*, *149*, 116-133. doi:10.1016/j.jecp.2016.01.013

- Lagattuta, K. H., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development, 85*, 659-674. doi:10.1111/cdev.12154
- Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia, 44*, 2139-2148. doi:10.1016/j.neuropsychologia.2005.10.013
- Lecce, S. & Bianco, F. (2017). Working memory predicts children's changes in theory of mind development during middle childhood: A training study. Manuscript submitted for publication.
- Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014). Training preschoolers on 1st-order-false-belief understanding: Transfer on advanced ToM skills and metamemory. *Child Development, 85*, 2404-2418. doi:10.1111/cdev.12267
- Lecce, S., Caputi, M., & Hughes, C. (2011). Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement?. *Journal of Experimental Child Psychology, 110*, 313-331. doi:10.1016/j.jecp.2011.04.011
- Lecce S., Caputi M., & Pagnin, A. (2014). Long-term effect of theory of mind on academic achievement: The role of sensitivity to criticism. *European Journal of Developmental Psychology, 11*, 305-318. doi:10.1080/17405629.2013.821944
- Lecce, S., Zocchi, S., Pagnin, A., Palladino, P., & Taumoepeau, M. (2010). Reading minds: The relation between children's mental state knowledge and their metaknowledge about reading. *Child Development, 81*, 1876-1893. doi:10.1111/j.1467-8624.2010.01516.x
- Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and

theory of mind at 3, 4, and 5 years. *Cognitive Development*, 33, 40-55.

doi:10.1016/j.cogdev.2014.07.001

McAuley, T. & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology*, 108, 453–468. doi:10.1016/j.jecp.2010.08.009

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, 135, 749-773. doi:10.1037/a0016854

Miller, S. A. (2012). *Theory of mind: Beyond the preschool years*. New York, NY, USA: Psychology Press.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21, 8-14. doi:10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.
doi:10.1006/cogp.1999.0734

Murphy, B. C., Shepard, S. A., Eisenberg, N., & Fabes, R. A. (2004). Concurrent and across time prediction of young adolescents' social functioning: The role of emotionality and regulation. *Social Development*, 13, 56-86. doi:10.1111/j.1467-9507.2004.00257.x

Muthèn, L. K., & Muthèn, B. (2012). *Mplus: Statistical Analysis with Latent Variables User Guide*. (7th ed.). Los Angeles, CA: Muthèn and Muthèn

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. New York, NY, US: Routledge/Taylor & Francis Group.

Orsini, A. (1997). *WISC-R. Contributo alla taratura italiana* (Contribution to Italian norms). Firenze, Italy: Organizzazioni Speciali.

Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks.

Child Development, 87, 1971-1991. doi:10.1111/cdev.12566

Oswald, T. M. (2013). Relations among theory of mind and executive function abilities in typically developing adolescents and adolescents with asperger's syndrome and high functioning autism (Doctoral dissertation). Retrieved from

<https://scholarsbank.uoregon.edu/xmlui/handle/1794/12529>

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. doi:10.1111/1467-8624.00436

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437-471. doi:10.1016/0022-0965(85)90051-7

Psychology, 39, 437-471. doi:10.1016/0022-0965(85)90051-7

Petermann, F., & Petermann, U. (Eds.). (2007). *Hamburg-Wechsler-Intelligenztest für KinderIV (HAWIK-IV)*. Göttingen: Hogrefe.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*, 127-152. doi:10.1080/17405620344000022

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011).

Development of hot and cool executive function during the transition to adolescence.

Journal of Experimental Child Psychology, 108, 621-637.

doi:10.1016/j.jecp.2010.09.008

Rice, K., Anderson, L. C., Velnoskey, K., Thompson, J. C., & Redcay, E. (2016). Biological motion perception links diverse facets of theory of mind during middle childhood.

Journal of Experimental Child Psychology, 146, 238-246.

doi:http://dx.doi.org/10.1016/j.jecp.2015.09.003

Roach, H. (Producer), Newmeyer, F. C. (Director), Taylor, S. (Director). (1923). *Safety last!* [motion picture]. Culver City, CA: Hal Roach Studios.

Ronald, A., Viding, E., Happé, F., & Plomin, R. (2006). Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: A twin study. *Social Neuroscience, 1*, 412-425. doi:10.1080/17470910601068088

Röthlisberger, M., Neuenschwander, R., Michel, E., & Roebbers, C. M. (2010). Exekutive Funktionen: Zugrundeliegende kognitive Prozesse und deren Korrelate bei Kindern im späten Vorschulalter. *Z. Entwicklungspsychol. Pädagog. Psychol., 42*, 99–110. doi:10.1026/0049-8637/a000010

Rubini, V. & Rossi, M. A. (1982). Analysis of the psycho-metric characteristics of the Primary Mental Abilities Tests (Level 4-sup-6, 1962 revision) applied to a group of Italian subjects. *Bollettino di Psicologia Applicata, 161–164*, 87–105. Available from: PsycINFO, Ipswich, MA. Accessed September 18, 2014.

Russell, J. (1996). *Agency: Its role in mental development*. Hove, UK: Taylor & Francis.

Sartorra, A. & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test staistic. *Psychometrika, 75*, 243-248. doi:10.1007/s11336-009-9135-y

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 298*, 199–209.

doi:10.1098/rstb.1982.0082

Shing, Y. L, Lindenberger U., Diamond, A., Li, S. C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence.

Developmental Neuropsychology 35, 679–697. doi:10.1080/87565641.2010.508546

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scales*.

Chicago, IL: Riverside.Thurstone, L. L., & Thurstone, T. G. (1962). *Primary mental abilities*. Chicago, USA: Science Research.

- Van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427–449. doi:10.1016/j.intell.2006.09.001
- Vetter, N. M., Altgassen, M., Phillips, L., Mahy, C. E. V., & Kliegel, M. (2013). Development of affective theory of mind across adolescence: Disentangling the role of executive functions. *Developmental Neuropsychology, 38*, 114-125. doi:10.1080/87565641.2012.733786
- Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a non verbal task. *Neuroimage, 29*, 90–98. doi:10.1016/j.neuroimage.2005.07.022
- Wang, Z., Devine, R. T., Wong, K., & Hughes, C. (2016). Theory of mind and executive function during middle childhood across cultures. *Journal of Experimental Child Psychology, 149*, 6-22. doi:10.1016/j.jecp.2015.09.028
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541. doi:10.1111/j.1467-8624.2004.00691.x
- White, S., Hill, E., Happé, F. G. E., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalising impairments in Autism. *Child Development, 80*, 1097-1117. doi:10.1111/j.1467-8624.2009.01319.x
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575-587. doi:10.1037/0012-1649.44.2.575
- Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PloS One, 8*, 1-9. doi:10.1371/journal.pone.0077770

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). Development of executive function in childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.

Figure captions

Figure 1. *Unstandardized and Standardized Parameter Estimates for Longitudinal Structural Equation Model*

Appendix A

Example Item and Scoring Criteria from the Strange Stories Task (White et al., 2009)

Jill wanted to buy a kitten, so she went to see Mrs. Smith, who had lots of kittens she didn't want. Now Mrs. Smith loved the kittens, and she wouldn't do anything to harm them, though she couldn't keep them all herself. When Jill visited she wasn't sure she wanted one of Mrs. Smith's kittens, since they were all males and she had wanted a female. But Mrs. Smith said, "If no one buys the kittens I'll just have to drown them!"

QUESTION: Why did Mrs. Smith say that?

2 points—reference to persuasion, manipulating feelings, trying to induce guilt/pity

1 point—reference to outcome (to sell them or get rid of them in a way which implies not drowning) or simple motivation (to make Jill sad)

0 points—reference to general knowledge without realization that the statement was not true (she's a horrible woman)

Appendix B

Example Item and Scoring Criteria from the Silent Film Task (Devine & Hughes, 2013)



Note. Screenshots taken with permission. From ‘Safety Last’ (Copyright of the Harold Lloyd Estate, 1923).

QUESTION: Why did the driver lock Harold in the van?

2 points—The driver didn’t know Harold was in the van; He didn’t mean/intend to.

1 point—He wanted/had to continue on his rounds (factual); He did not see/hear him (no explicit reference to a mental state).

0 point—The man is deaf/hard of hearing (no mention of ‘knowing’); He was reading on the van; He wasn’t supposed to be there; The man told him to; He kidnapped him.

Note. From “Silent Films and Strange Stories: Theory of mind, gender and social experiences in middle childhood” by Devine & Hughes (2013). *Child Development*, 84, p.993.

Appendix C

Preliminary Analysis

Correlations between Strange Stories and Silent Film tasks were significant within each time point of the study, $.24 < r < .30$, all $ps < .01$. These correlations were weaker than those reported in previous studies using error-free latent variable correlations (Devine & Hughes, 2013; 2016) but consistent with those reporting the correlations between raw summed scores (Devine et al., 2016). A confirmatory factor analysis (CFA), using a robust maximum likelihood estimator, in which each of the two theory-of-mind tasks were permitted to load onto a single latent variable at each time point provided an acceptable fit to the data, $\chi^2(1) = 0.24$, $p = 0.63$, CFI = 1.00, TLI = 1.09, RMSEA = 0. At each time-point both theory-of-mind tasks loaded significantly onto the respective latent factor, all standardized loadings $>.34$, all $ps <.001$. Moreover, the factor determinacy co-efficients were .83 at Time 1, .85 at Time 2, and .88 at Time 3, indicating high internal consistency (Brown, 2015).

Appendix D

Comparison of Nested Models

We examined whether placing equality constraints on specific model parameters would result in changes of fit relative to our baseline (comparison) model (reported above). Since we used a robust maximum likelihood estimator with a mean-adjusted χ^2 test statistic, we calculated the χ^2 difference between each nested model and the comparison model using the Satorra-Bentler χ^2 difference test (Satorra & Bentler, 2010). Model fit statistics for each of these nested models are presented in Table 3.

Table 3

Model Comparisons

Model	CFI	TLI	RMSEA [90% CI]	χ^2	df	$\Delta\chi^2$
Baseline (Comparison Model): No constraints	1.00	1.01	0 [0, 0.07]	29.50	31	-
Nested Model 1: T1WM→T2ToM = T1ToM→T2WM	0.99	0.99	0.02 [0, 0.07]	33.87	32	3.39
Nested Model 2: T2WM→T3ToM = T2ToM→T3WM	0.98	0.96	0.04 [0, 0.09]	38.97	32	8.14**
Nested Model 3: T1IC→T2ToM = T1ToM→T2IC	0.99	0.97	0.04 [0, 0.08]	36.94	32	6.06*
Nested Model 4: T2IC→T3ToM = T2ToM→T3IC	0.99	0.99	0.03 [0, 0.08]	34.42	32	3.81
Nested Model 5: T1IC→T2ToM = T1WM→T2ToM	0.98	0.96	0.04 [0, 0.09]	38.92	32	6.74**
Nested Model 6: T2IC→T3ToM = T2WM→T3ToM	0.98	0.96	0.05 [0, 0.09]	39.77	32	6.75**

Note. ** $p < .01$. * $p < .05$. CFI = Cumulative Fit Index. TLI = Tucker Lewis Index. RMSEA = Root Mean Square Error of Approximation. T1 = Time 1. T2 = Time 2. T3 = Time 3. WM = Working Memory. IC = Inhibition. ToM = Theory of Mind. Arrows depict longitudinal regression paths. $\Delta\chi^2$ = Sartorra-Bentler scaled chi-square difference.

First, to examine the asymmetry in the longitudinal association between WM and ToM, we constrained the cross-lagged paths from Time 1 to Time 2 (Nested Model 1) and from Time 2 to Time 3 (Nested Model 2) to equality. Nested Model 1 did not produce a statistically significant decrease in model fit relative to the baseline model indicating that there was no significant difference in the strength of these two paths, $\Delta\chi^2(1) = 3.39, p = .07$. This means that the path from ToM at Time 1 to WM at Time 2 did not differ in strength from the path from WM at Time 1 to ToM at Time 2. In contrast, there was a significant decrease in model fit between Nested Model 2 and the baseline model, $\Delta\chi^2(1) = 8.14, p = .004$, indicating that there was a significant difference between two regression paths. That is, there was a unidirectional longitudinal association between Time 2 WM and Time 3 ToM but not vice versa.

Second, to examine the cross-lagged longitudinal associations between IC and later ToM, we estimated two further nested models in which we constrained the cross-lagged paths from Time 1 to Time 2 (Nested Model 3) and from Time 2 to Time 3 (Nested Model 4) to equality. Nested Model 3 produced a significant decrease in model fit relative to the baseline model, $\Delta\chi^2(1) = 6.06, p = .013$, indicating that the path from Time 1 IC to Time 2 ToM differed in strength from the path from Time 1 ToM to Time 2 IC. Specifically, there was a unidirectional negative association between Time 1 IC and Time 2 ToM but not vice versa. Nested Model 4 produced a slight but statistically non-significant decrease in model fit relative to baseline, $\Delta\chi^2(1) = 3.81, p = .06$, indicating that there was no significant difference in the strength of the regression paths linking Time 2 IC with Time 3 ToM and linking Time 2 ToM with Time 3 IC. This means that there were no significant longitudinal associations between either construct from Time 2 to Time 3 and that the strength of these cross-lagged paths did not differ from one another.

Third, to compare the strength of the longitudinal associations between WM and ToM and between IC and ToM, we constrained the cross-lagged paths from Time 1 WM to Time 2 ToM and Time 1 IC to Time 2 ToM to equality (Nested Model 5). We then constrained the paths from Time 2 WM to Time 3 ToM and from Time 2 IC to Time 3 ToM to equality (Nested Model 6). Nested Model 5 resulted in a significant decrease in model fit relative to the baseline model, $\Delta\chi^2(1) = 6.74, p = .009$, indicating that there was a significant difference in the strength of the longitudinal associations between Time 1 WM and Time 2 ToM and between Time 1 IC and Time 2 ToM. These paths differed significantly in strength from one another with Time 1 WM showing a positive association with Time 2 ToM and Time 1 IC showing a negative association with Time 2 ToM. There was also a significant decrease in model fit between Nested Model 6 and the baseline model, $\Delta\chi^2(1) = 6.75, p = .009$, indicating that there was a significant difference in the strength of the association between Time 2 WM and Time 3 ToM and between Time 2 IC and Time 3 ToM. That is, while Time 2 WM was a significant predictor of Time 3 ToM, IC at Time 2 did not predict Time 3 ToM.