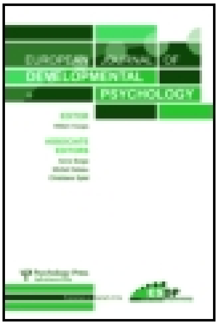

European Journal of Developmental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pedp20

Dynamics of the Theory of Mind construct: A developmental perspective

Mieke P. Ketelaars [a], Marjolijn van Weerdenburg [a], Ludo Verhoeven [a], Juliane M. Cuperus [b] & Kino Jansonius [b]

[a] Radboud University Nijmegen, Behavioural Science Institute, Nijmegen, The Netherlands

[b] Sint Marie, Eindhoven, The Netherlands

Version of record first published: 08 Jan 2010

To cite this article: Mieke P. Ketelaars, Marjolijn van Weerdenburg, Ludo Verhoeven, Juliane M. Cuperus & Kino Jansonius (2010): Dynamics of the Theory of Mind construct: A developmental perspective, European Journal of Developmental Psychology, 7:1, 85-103

To link to this article: http://dx.doi.org/10.1080/17405620903482081

EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY
2010, 7 (1), 85–103

Psychology Press
Taylor & Francis Group

Dynamics of the Theory of Mind construct: A developmental perspective

Mieke P. Ketelaars, Marjolijn van Weerdenburg and Ludo Verhoeven
Radboud University Nijmegen, Behavioural Science Institute, Nijmegen, The Netherlands

Juliane M. Cuperus and Kino Jansonius
Sint Marie, Eindhoven, The Netherlands

Theory of Mind (ToM) encompasses a wide variety of abilities, which develop during childhood. However, to date most ToM research has focused on the single concept of false-belief understanding, and examined ToM only in young children. Furthermore, there is a lack of implementation of a longitudinal design, which examines the dynamics of the ToM construct over several years. Our longitudinal study measured the abilities of a group of 5-year-old children ($n = 77$) in mainstream education during three consecutive years, on aspects of ToM related to emotion understanding and false-belief understanding. The results provide support for significant improvements in emotion understanding and false-belief understanding between the ages of 5 and 7. Whereas emotion attribution was already largely developed at age 5, more intricate aspects of emotion understanding, such as understanding display rules and understanding mixed emotions showed significant developments. Over the course of the years, children also showed an increasing awareness of false-belief understanding. In addition to the developmental growth, the different ToM aspects were found to be relatively stable over time. Correlations as well as predictive relations between emotion understanding and false-belief understanding could be identified. Finally, there was evidence for the role of language ability in the development of the ToM aspects under consideration. The results support the notion that ToM abilities measured at age 5 are not just a snapshot but provide a longer-term outlook as well.

Keywords: Development; Emotion understanding; False-belief understanding.

Correspondence should be addressed to Mieke Ketelaars, Radboud University Nijmegen, PO Box 9104, NL-6500 HE Nijmegen, The Netherlands. E-mail: M.Ketelaars@pwo.ru.nl

http://www.psypress.com/edp DOI: 10.1080/17405620903482081

The ability to understand other people's mental states, known as Theory of Mind (ToM), is an important marker in child development, and has been studied extensively (e.g., Baron-Cohen, 2001; Bartsch, 1998; Cutting & Dunn, 1999; Flavell, Flavell, & Green, 1990; Hughes et al., 2000; Moses & Flavell, 1990; Perner & Lang, 1999; Repacholi & Slaughter, 2003; Wellman & Liu, 2004). Whereas early studies on ToM often applied a single-task paradigm limited to false-belief understanding (see Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004, for an overview), more recently researchers have started to embrace the wide variety of abilities that are connected to ToM (e.g., Cutting & Dunn, 1999; Wellman & Liu, 2004), which also includes several aspects of emotion understanding. This has led to the view that early ToM abilities may function as stepping stones for later abilities. However, although there has been an increase in awareness of developmental issues, studies adopting a longitudinal design are still scarce (but see Flynn, 2006, for an example). Those that do apply a longitudinal design often investigate only limited aspects of ToM. For instance, Amsterlaw and Wellman (2006) investigated the gradual development of false-belief understanding, as did Flynn (2006). Brown and Dunn (1996), and Pons and Harris (2005) studied the development of several aspects of emotion understanding. Yet we know little of the interactions involved in the development of different aspects of ToM. This study attempts to answer some issues related to this gap using a large sample of children who have been studied longitudinally over a period of two years. Before going into the design and results of the present study, a short overview will be given of the key ToM concepts of emotion understanding and false-belief understanding, and their developmental trajectories.

According to Buitelaar, Van der Wees, Swaab-Barneveld, and Van der Gaag (1999): "emotion understanding includes the ability to discriminate the various expressions of emotions in facial, gestural, and verbal display and, in doing so, to be sensitive to dimensions of intensity, complexity and contextual colouring" (p. 869). Basic levels of emotion understanding consist of visual recognition of emotions and their causes. As Brown and Dunn (1996) demonstrated, children as young as three years have been reported to understand causal attributions of emotions. The more intricate aspects of emotion understanding do not develop until the early and middle school years (Brown & Dunn, 1996).

One of the more advanced levels of emotion understanding consists of the ability to understand that different (conflicting) emotions can occur together (Brown & Dunn, 1996). Understanding these so-called mixed emotions is manifested for example when children realize they can be both afraid and excited on a ride in a roller-coaster (Kestenbaum & Gelman, 1995). It has been stated that children under the age of 7 appear to experience problems understanding mixed feelings, although they do seem to be able to express their own ambivalent feelings from an early age (Kestenbaum & Gelman, 1995).

A second advanced level of emotion understanding consists of the ability to understand and apply display rules that govern emotions (Underwood, Coie, & Herbsman, 1992). At a young age, children begin to understand that there are social norms that apply in the expression of emotion (Banerjee, 1997). For example, they conceal their real emotion upon receiving a present they do not really like. Consistent with the developmental pathway of mixed emotions, Banerjee (1997) found the application of display rules to develop before the understanding of these rules. Moreover, it seems that this ability gradually develops during childhood: in a study by Saarni (1979) only 10-year-old children were able to offer a justification when asked about display rules, but 6- and 8-year-olds did understand rules governing emotions when prompted specifically. Banerjee (1997), however, found that children as young as 3 years were able to understand the intricacies involved in display rules.

Emotion understanding seems to evolve gradually and in several steps. In a longitudinal design Pons and Harris (2005) found that emotion understanding, as measured with a variety of tasks, followed a predictable course of change in children aged 7 to 11 through three stages. At the first stage children understand external cues of emotions, e.g., facial expressions, whereas at the second stage they also understand mental aspects of emotions, e.g., the role of beliefs and desires on emotions. The final stage consists of more reflective aspects of emotion, e.g., mixed emotions. Pons and Harris also found a certain stability of emotion understanding, i.e., children with low emotion understanding showed similar low performance one year later. In an older study by Hughes and Dunn (1998), longitudinal correlations in emotion understanding were also visible in a younger sample of 4- and 5-year-old children over a period of one year.

The second key concept of ToM consists of false-belief understanding. False-belief understanding is defined as the ability to understand that others can have an inaccurate understanding (a false belief) of reality. As stated, false-belief understanding is often considered to be the litmus test of ToM (Wellman et al., 2001). In a comprehensive meta-analysis on the emergence of first-order false-belief understanding, Wellman et al. (2001) found evidence for a "conceptual shift" between the ages of 2½ and 5 years. A second developmental shift occurs at around the age of 6 or 7, when children are also able to succeed on second-order false-belief tasks, i.e., beliefs about beliefs (Perner & Wimmer, 1985). Nevertheless, as shown by Mutter, Alcorn, and Welsh (2006), first-order false-belief understanding is by no means fully developed by the age of 5, and great individual variance exists in the age at which children are able to successfully pass tests of false belief. As with emotion understanding, false-belief understanding develops gradually, with clear precursors such as understanding diverse beliefs and desires (Wellman & Liu, 2004). Although false-belief understanding is often considered a hallmark in the cognitive development of ToM, the longitudinal stability of false-belief understanding has not been established unequivocally. For instance, Hughes and Dunn (1998) could not establish a significant correlation in ToM performance of 4-year-olds and their performance one year later, although they did find several significant correlations over a period of half a year. Similarly, in another study by Hughes (1998), no significant correlation was found over a period of one year.

Emotion understanding and false-belief understanding are related in nature, since they both hinge on an understanding of subjective mental states rather than reality and thus require reasoning about beliefs and desires (Bloom, 2003). Despite this relation, research in these areas has been developing quite independently (Cutting & Dunn, 1999). Buitelaar et al. (1999) found a moderate correlation between false-belief understanding and emotion understanding. However, their research was based solely on emotion recognition using visual expressions. Converging evidence comes from a study by Cutting and Dunn (1999), who showed that both aspects should be considered distinct abilities, though they are significantly related (with correlations between .39 and .44). Their conclusion was based on the fact that neither variable contributed unique variance once family background, age, and language were taken into account. De Rosnay, Pons, Harris, and Morrell (2004) also found a significant correlation between emotion attribution and false-belief understanding. It should be noted that they assessed the ability to make (false) belief-based emotion attributions.

In conclusion, longitudinal research into ToM viewed as a dynamic construct involving a variety of skills is scarce. Although a start has been made in studying ToM in a longitudinal setting, the main bulk of literature has focused on cross-sectional data. In addition, the studies that have used a longitudinal design have focused mainly on young children. This is not without reason, since major steps in the development of ToM take place during these early years. However, ToM is by no means fully developed by this time. It would, therefore, be worthwhile to investigate the developmental trajectories of ToM beyond the early preschool years. This study attempted to address some issues regarding these later developments. More specifically, this study had the following goals:

1. To examine the developmental progression of ToM.
2. To gain more insight into the dimensions of ToM and their stability over time.
3. To explore the dynamics of ToM abilities as a function of children's age.
4. To explore the dynamics of ToM abilities as a function of children's language ability.

Additionally, we considered the influence of gender and socioeconomic status (SES) of the parents in relation to the developmental progression of ToM.

METHOD

Participants

The sample included 77 children (54 boys, 23 girls) who were part of a larger study. All children spoke Dutch as a first language and did not show any learning problems or behavioural problems according to their teachers. Their mean age at Time 1 was 5;6 years, at Time 2 their mean age was 6;5 years and at Time 3 their mean age was 7;5 years. Additional information on the participants is provided in Table 1. Teachers were also asked to classify the highest level of completed education of the parents on a 4-point scale. A score of 1 was given when parents finished only elementary education. A score of 2 was given when parents had a degree in lower general secondary education, whereas a score of 3 was given when parents had a degree in higher general secondary education. A score of 4 was given when parents attained a college or university degree. No parents were classified into the lowest educational level.

Measures

For the measurement of ToM abilities a booklet was created by the authors. The booklet contained pictures that were accompanied by stories. All items in the tasks consisted of similar procedures using verbal stories and visual aid (see Figure 1 for a sample of a story illustration). The faces of the story characters were depicted in a neutral fashion. During the stories, several questions were asked regarding emotion understanding (emotion attribution, mixed emotions, display rules) and false-belief understanding (change of location, conceptual perspective taking). The following paragraphs provide information on the tasks and the scoring of the different ToM tasks.

TABLE 1
Participants information

	Time 1	*Time 2*	*Time 3*
Number of participants	77	76	74
Boys/girls	54/23	54/22	52/22
Age in years (*SD* in months)	5;6 (3.5)	6;5 (3.3)	7;5 (3.8)
Range of age (in months)	59–73	72–86	83–99

Figure 1. Story illustration sample.

Emotion attribution. The children were presented with four emotion-attribution tasks. In these tasks, children heard an emotion-evoking story (e.g., finding out someone is cheating while playing hide and seek). Having heard the story, children were asked to make an emotion attribution: "*How would [story character] feel? Would he/she feel happy, sad, angry or scared?*" Children received a score of 2 for an item when the answer was considered correct. A score of 1 was granted when a child substituted an emotion with a correctly poled emotion (either negative or positive) that was considered partly correct but simplified (i.e., substituting the emotion attribution scared for the emotion sad and angry when someone has run away from home). Since one of the items was considered too easy, with a success rate of 97% at Time 1, it was decided to eliminate this item from the analyses. The score for this task is the mean score achieved on the three items.

Mixed emotions. The children were presented with two mixed-emotion tasks. In these tasks children heard a story that would provoke two conflicting emotions in the story character (e.g., two friends are playing soccer in a match. One of them scores, while the other one does not although he really wanted to.) Having heard the story, children were asked an emotion-attribution question and a justification question: "*Would [story character] feel happy, sad or both happy and sad at the same time? Why?*" The justification question had to be answered correctly to get credited a score of 2. In all other cases a score of 0 was assigned. The score for this task is the mean score achieved on the two items.

Display rules. The children were presented with two display-rules tasks. In these tasks children heard an emotion-evoking story in which the hiding of the emotion would be provoked (e.g., being scared to go off a high slide in the swimming pool, but having a friend nearby who does want to go off the slide). Having heard the story, children were asked whether the protagonist would display the emotion or not. Subsequently they were asked to justify their answer: "*Would [story character] show that he/she is scared? Why?*" A probe question was asked if a child did not answer correctly. The probe question was asked to inform the child that a choice existed: "*In this story [story character] can either choose to show her feelings or hide them. What would you do? Why?*" Answers were only considered correct and granted a score of 2 when either one of the questions was answered correctly in combination with an accurate justification. In all other cases a score of 0 was assigned. The score for this task is the mean score achieved on the two items.

False-belief understanding. The children were presented with three change-of-location, first-order false-belief tasks (e.g., a boy hiding from his mother after playing in the living room—subsequently the mother tries to find her son). Change of location could include a person as in the example, but also an object. After hearing the story, children were asked about the false belief of one of the story characters: "*Where would [story character] look for [the object]?*" To obtain a score of 1 on the first-order false-belief questions, a memory question pertaining to the actual location of the object had to be answered correctly as well as the false-belief question. The score for this task is the mean score achieved on the three items.

Conceptual perspective taking. The children were presented with two conceptual-perspective-taking tasks. Adapted from Lalonde and Chandler (2002), children were shown a picture and asked what was on it. Subsequently, part of the picture was covered, in such a way that only a small part (which could not be identified as the whole picture) was still visible. Children were asked to infer about the belief of a classmate: "*If a classmate walked into the room and saw this picture, what would he/she think was on it?*" Only answers not pertaining to the true nature of the picture were considered correct and granted a score of 1. Following the scoring procedure of Lalonde and Chandler, answers that referred partly to the true nature of the picture were granted a score of 1. The score for this task is the mean score achieved on the two items.

Phi correlations for each of the items within a subtask were computed. All items within a subtask correlated significantly ($p < .05$), except for the phi correlation between two emotion understanding items ($p = .07$) and the phi correlation between two false-belief understanding items.

Receptive vocabulary. For measuring receptive vocabulary, a subtest of the Dutch Language Test for Children (Verhoeven & Vermeer, 2001) was administered. The TAK is a standardized test for 4- to 10-year-old children. In the receptive vocabulary subtest, the child is presented with a word and is asked to select the picture illustrating that word out of four pictures. The maximum number of items is 96, and the task is discontinued after five consecutive errors.

Procedure

At all points in time, the children were tested at their schools in two sessions of approximately 50 minutes each. Upon entering the room the children were first familiarized with the situation and with the experimenter. Since the ToM items were asked over the course of several stories, the tasks were completed in a fixed order. The measurements at Time 1, Time 2, and Time 3 were separated by a time interval of approximately one year.

RESULTS

Developmental progression of ToM

To assess the developmental progression of ToM abilities, means and standard deviations were computed for all ToM tasks at Time 1, Time 2, and Time 3. Next, repeated-measures analyses of variance (ANOVAs) with planned repeated contrasts were performed to assess whether there was significant growth in ToM abilities from Time 1 to Time 2, and from Time 2 to Time 3. As can be gathered from Table 2, the repeated-measurement ANOVAs showed a time effect for four of the five ToM tasks.

The repeated-measures ANOVA on the emotion-attribution scores was the only task that did not reveal a significant time effect. No increase in scores was visible between either time point. All other increases were significant with the following exceptions: the display rules score did not

TABLE 2
Means and standard deviations of ToM scores at Time 1, Time 2, and Time 3 and the results of the repeated-measures ANOVA

	Time 1		*Time 2*		*Time 2*					
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*df*	*p*	η_p^2
Emotion attribution	1.54	0.35	1.55	0.37	1.56	0.35	0.13	2,142	.88	.00
Mixed emotions	0.46	0.65	0.86	0.81	1.21	0.84	22.37	2,142	.00	.24
Display rules	0.16	0.43	0.63	0.66	0.74	0.73	10.51	2,84	.00	.20
False belief	0.61	0.35	0.72	0.28	0.81	0.26	14.11	2,142	.00	.17
Conceptual perspective taking	0.95	0.84	1.52	0.71	1.69	0.58	24.11	2,118	.00	.29

Note: Range of Emotion attribution scores = 0–2; Range of Mixed emotions score = 0–2; Range of Display rules score = 0–2; Range of False belief score = 0–1; Range of Conceptual perspective score = 0–2.

increase from Time 2 to Time 3, and neither did the conceptual perspective-taking score.

Of the five ToM tasks standardized *z*-scores were computed with $M = 0$ and $SD = 1$ using the entire 3-year set of measurements. Figure 2 graphically depicts the development of the different ToM abilities over time, using the *z*-scores to allow for comparisons between the tasks.

Although we found evidence for developmental growth on several measures, we wanted to roughly assess the year that children started to grasp the notion of the tasks. For this, one-sample *t*-tests were performed. An arbitrary crossover performance level of 50% correct answers was chosen to indicate an awareness of the measured ability. A one-sample *t*-test on the mixed-emotions task showed that the scores at Time 1 fell significantly below the 50% performance level ($p < .001$), indicating a disposition expecting the sensation of one emotion rather than two. At Time 2 the mean score on the mixed-emotion task did not differ from our reference performance level ($p = .27$). Similarly, the mean performance on the conceptual-perspective-taking task at Time 1 did not differ from our reference performance level ($p = .55$). Finally, the mean performance on the display-rules task did not reach our reference performance level at any time point ($p < .01$ for Time 1 and Time 2, $p < .05$ for Time 3), indicating that the understanding of display rules is still largely unknown territory for our 7-year-olds.

Effect of socioeconomic status and gender

To explore the influence of parental educational level (both paternal and maternal) on ToM performance Pearson's correlations were computed. The

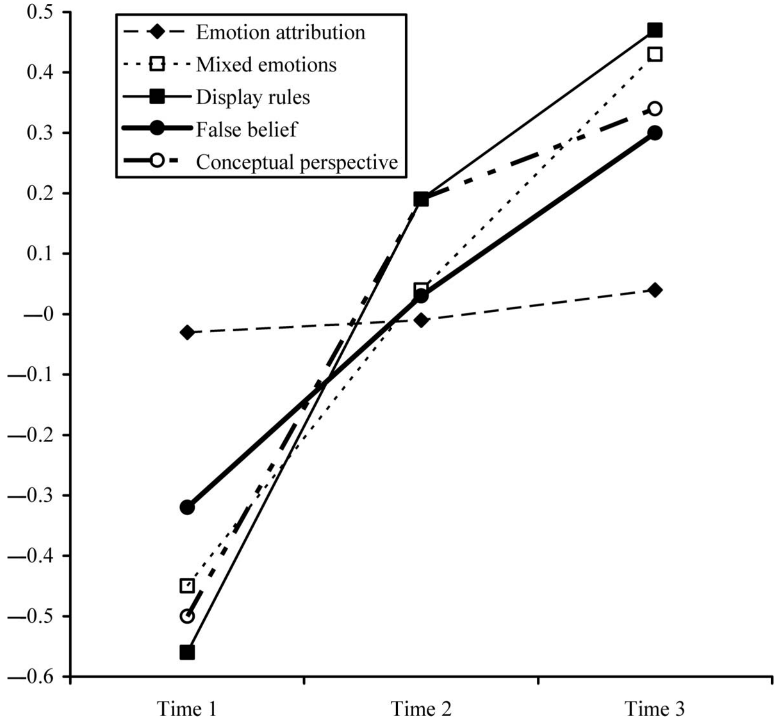

Figure 2. Developmental progression of ToM abilities over Time 1, Time 2, and Time 3.

paternal educational level correlated with ToM performance to a limited extent. It was only correlated to the first-order false-belief score at Time 3 ($r = .30$, $p < .05$). Maternal educational level, however, persistently correlated with several ToM aspects. At Time 1, it was positively correlated with the false-belief score ($r = .32$, $p = .01$). The correlation with false-belief understanding persisted at Time 2 ($r = .22$, $p = .08$, marginally significant) and Time 3 ($r = .35$, $p < .01$). Moreover, maternal educational level was also persistently correlated with the understanding of mixed emotions ($r = .29$, $p < .05$; $r = .27$, $p < .05$; and $r = .29$, $p < .02$, for the three consecutive time points). In addition there was a significant negative correlation at Time 3 with the emotion-attribution task ($r = -.29$, $p < .05$). This result can possibly be explained by the fact that the relatively easy emotion-attribution task put some children on the wrong track: they came up with more intricate emotions, and thus failed the task. To explore possible gender differences in ToM understanding, we conducted *t*-tests. No significant differences were found for gender.

Dimensions of ToM

To verify that our ToM items could indeed be subdivided into several aspects of ToM abilities, we performed a principal component analysis with varimax rotation at all time points (see Table 3). Since even our 7-year-olds did not show an active awareness on our display-rules task at any time point, it was decided to eliminate this task from the analysis. The factor analysis provided four factors which explained 59% of the variance. The first factor was related to both conceptual perspective-taking items and explained roughly 22% of the variance. The second factor was related to the first-order false-belief items and explained roughly 14% of the variance. The third factor consisted of the mixed-emotion items, explaining 13% of the variance. The final factor consisted of the emotion-attribution items and explained an additional 10% of the variance. As such, the results of the factor analysis supported our division into four sets of ToM-related tasks.

Stability in change of ToM over time

Using structural equation modelling (SEM) in AMOS 6.0 (Arbuckle, 2005), longitudinal stability of the various ToM abilities was investigated for each of the measured ToM abilities using quasi-simplex models. The goodness-of-fit statistics of the four models are shown in Table 4 (Models A–D). Goodness of fit of the models was assessed by several indices: the standard χ^2 test and alternative goodness-of-fit indices such as the *Adjusted Goodness of Fit Index* (AGFI), the *Comparative Fit Index* (CFI), the *Normed Fit Index* (NFI), the *Root Mean Square Error of Approximation* (RMSEA) and the *Standardized Root Mean Square Residual* (SRMR). The AGFI, CFI, and

TABLE 3
Varimax rotated four-factor solution for the ToM items on all time points combined

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*
Conceptual perspective 1	.83			
Conceptual perspective 2	.89			
Emotion attribution 1				.66
Emotion attribution 2				.58
Emotion attribution 3				.62
False belief 1		.61		
False belief 2		.75		
False belief 3		.63		
Mixed emotions 1			.71	
Mixed emotions 2			.86	

Note: Eigenvalues > 1.0; Values > .30 are reported.

TABLE 4
Goodness of fit statistics for structural models at three time points

Model	χ^2	*df*	*p*	*AGFI*	*CFI*	*NFI*	*RMSEA*	*SRMR*
A: Quasi-simplex emotion attribution	7.02	1	.01	.67	.00	.17	.28	.12
B: Quasi-simplex mixed emotions	2.36	1	.12	.88	.92	.88	.13	.06
C: Quasi-simplex false belief	4.36	1	.04	.79	.90	.88	.21	.08
D: Quasi-simplex conceptual perspective	2.68	1	.10	.87	.90	.88	.15	.07
E: Final integrated model	28.07	24	.26	.87	.96	.78	.05	.09

NFI should ideally be higher than .80 (Hu & Bentler, 1999). The RMSEA should be lower than .05 to reflect a good fit, while values of .08 indicate a reasonable fit (Browne & Cudeck, 1993). The SRMR finally should ideally be below .08 to reflect a good fit (Hu & Bentler, 1999), while values below .10 are considered acceptable.

As the different goodness-of-fit indices suggest, the fit of most quasi-simplex models was acceptable, with the exception of the emotion-attribution measure. This exception can be explained by the presence of a ceiling effect: due to the high scores of children at Time 1, any skill improvements of these children at later times are not registered, while decreases in skill levels are. Due to this anomalous effect, we investigated the stability of emotion attribution using an alternative technique. We analysed the percentages of children who regressed, performed at a similar level, and showed improved performance. Of all children who did not achieve the maximum score at Time 1 ($n = 66$), 82% performed at a similar level or improved at Time 2. The results from Time 2 to Time 3 were similar.

Dynamics of ToM over time

After establishing the stability of three of the four factors, SEM analyses were conducted to investigate the dynamic relations between the different ToM abilities. Since the emotion-attribution tasks did not show an acceptable stability over time, this measure was not included. Applying an iterative process on the three quasi-simplex models, significant relationships were added to the model, to find an integrated model with the best fit. The fit indices of this integrated model (Model E) are presented in Table 4. As can be gathered from the indices, this integrated model shows a good fit. Figure 3 provides a graphical presentation of the integrated model. Only significant relations are shown in Figure 3, with the exception of the relationship between the understanding of mixed emotions at Time 1 and

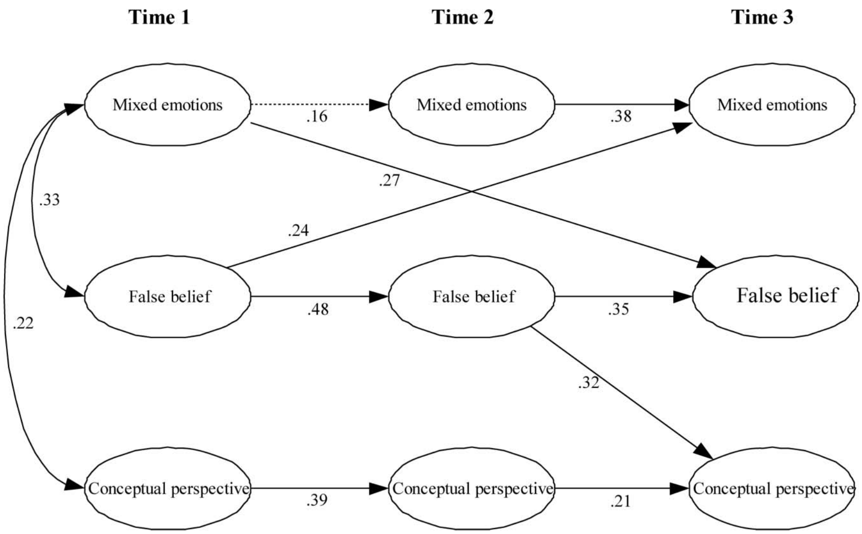

Figure 3. Final SEM model of relationships between ToM measures at Time 1, Time 2 and Time 3.

Time 2 (standardized regression coefficient = .16, $p = .15$). The figure shows that there are several relations between ToM abilities over time. At Time 1, there is a significant relation between the mixed-emotions score and the false-belief score. Additionally mixed-emotions understanding is also related to conceptual perspective taking. Longitudinally, several relations proved significant. For example, understanding mixed emotions at Time 1 predicts a modest but significant part of the performance of false-belief understanding at Time 3. Likewise, performance of understanding mixed emotions at Time 3 is predicted by false-belief understanding at Time 1. Furthermore, false-belief understanding at Time 2 predicts performance on the conceptual perspective taking task at Time 3.

To find out whether language would play a mediating role in ToM understanding, a last SEM model was computed applying an iterative process, including receptive vocabulary at Time 1 as a predictor for all time points. Scores on the receptive vocabulary task at Time 1 ranged from 30 to 89 ($M = 64.90$, $SD = 10.78$). The model showed a good fit according to the different fit indices, $\chi^2(28) = 26.77$, $p = .53$, AGFI = .88, CFI = 1.00, NFI = .85, RMSEA = .00, and SRMR = .06. Receptive vocabulary was found to predict significant variance of ToM performance. We found significant correlations between receptive vocabulary and the three ToM measures at Time 1 ($r = .28$ for mixed emotions, $r = .25$ for false-belief understanding, and $r = .26$ for conceptual perspective taking, all ps $< .05$).

Additionally, receptive vocabulary predicted a considerable part of variance of mixed-emotions understanding at Time 2 (standard regression coefficient = .52) and conceptual perspective taking (standard regression coefficient = .23). Finally, the predictive power of receptive vocabulary on mixed-emotions understanding persisted at Time 3 (standard regression coefficient = .25). All existing relations of the basic integrated model held when receptive vocabulary was entered into the model.

DISCUSSION

This longitudinal study attempted to address some issues on the developmental trajectories of several ToM abilities. These issues concern the development of ToM, its underlying dimensions and the dynamics of ToM abilities as a function of age.

With respect to our first goal, i.e., the development of ToM aspects of emotion understanding and false-belief understanding, we found evidence of developmental growth on two of the three measured aspects of emotion understanding and on both measured aspects of false-belief understanding. The greatest improvements in performance were found between the ages of 5 and 6 years. Skills related to emotion attribution did not improve during this period, which could be the result of a high level of emotion-attribution understanding at an earlier age. We found clear evidence for an emerging understanding of mixed emotions between 5 and 7 years. Whereas the performance of the 5-year-olds did not reveal a significant appreciation of mixed emotions, some 6-year-olds did reveal an appreciation of mixed emotions, and at age 7 the children understood mixed emotions more often than not. Similarly to the mixed-emotion understanding at age 5, we found no evidence for an active understanding of display rules at age 5. Although there was a significant improvement in performance on the display rules task from age 5 to 6, children still showed a predisposition to incorrectly predict that protagonists would show rather than hide emotions. Concerning the development of false-belief understanding, developmental growth was also established. Although literature often states that children reach this hallmark at around the age of 4 (Wellman et al., 2001, for a review), it does seem to be the case that false-belief understanding is still under development in the age range under consideration. In addition to a standard measure of false-belief understanding (i.e., change of location), we also used an alternative task following Lalonde and Chandler (2002). Whereas the "standard" false-belief tasks clearly have a right and a wrong answer, the conceptual-perspective-taking task demands a certain degree of creativity, since there is no such thing as the right answer. Though linguistically easier, the task might be more difficult on a conceptual level, a hypothesis that is corroborated by the fact that both our study and the study done by Lalonde and Chandler (2002) show relatively low scores on the conceptual-perspective-taking task at age 5. However, the children in our study showed a rapid improvement on this task. Overall, the results concerning developmental change in emotion understanding and false-belief understanding are largely consistent with the literature (Brown & Dunn, 1996; Kestenbaum & Gelman, 1995; Mutter et al., 2006; Saarni, 1979), and provide evidence of a growing awareness of more complex aspects of ToM, such as understanding mixed emotions and display rules.

As expected, performance on ToM tasks was related to socioeconomic status (SES). Maternal educational level was found to be significantly correlated to several ToM abilities, a finding that replicates results from other studies (e.g., Cutting & Dunn, 1999). This effect of maternal educational level was visible in all years. With regard to gender, our study did not show a significant difference between boys and girls.

Our second goal pertained to the underlying dimensions of ToM and their stability over the years. Taking into account the development of the ToM abilities, we found a robust stability across several tasks. As such, early emotion understanding and false-belief understanding seem to be predictive of later emotion understanding and false-belief understanding. As an exception to this established stability, we did not find evidence for stability of emotion attribution over the years. However, this might be related to the high initial level over emotion attribution, which led to a ceiling effect on our task.

The third goal of this study pertained to the dynamics of ToM. Some studies have indicated a lack of coherence within ToM batteries (Carlson, Mandell, & Williams, 2004). In our study we found some evidence for interrelations between emotion understanding and false-belief understanding. At Time 1 we found several correlations between emotion understanding and false-belief understanding. Surprisingly, our first-order false-belief task did not show any relation with the conceptual-perspective-taking task, which is theoretically related to false-belief understanding. The absence of a correlation between both tasks might have been the result of the fact that performance on the conceptual-perspective-taking task was still at a low level at the first year. However, this was not true for the scores at Time 2. We did find evidence for a relation from Time 2 to Time 3.

Another interesting finding related to the coherence of ToM abilities, is the result that understanding of mixed emotions at age 5 predicted false-belief understanding at age 7. Alternatively, false-belief understanding at age 5 served as a predictor of awareness of mixed emotions at age 7. Awareness of mixed emotions requires an understanding of two (conflicting) beliefs, which is similar to the understanding of false-belief understanding.

Finally, to find out whether language ability was related to emotion understanding and false-belief understanding, we added a measure of receptive vocabulary to our model. Language ability was found to correlate with ToM performance and also to predict a considerable amount of variance in ToM performance in following years. Several studies such as the study by Astington and Jenkins (1999) have established strong predictive relations between language skills and ToM over time. Whereas some relation between language and ToM is to be expected, due to the linguistic nature of ToM tasks, one can also argue for a more fundamental relationship (Astington & Jenkins, 1999). Although our results corroborate the relationship between language and ToM, it is not possible to answer questions pertaining to the nature of the relationship.

The results of this study shed more light on the developmental trajectories and interrelations in the realm of ToM skills. However, some limitations are worth mentioning. First, the ToM skills we measured consisted of a limited set of items. As a consequence, floor and ceiling effects were easily attained. In addition, it would have been preferred to counterbalance the order of presentation across the children. To investigate the age of acquisition of more difficult ToM skills such as display-rules understanding, a broader age range would need to be studied. Furthermore, it would seem useful to apply a microgenetic approach. Such an approach has recently been adopted in ToM studies (e.g., Amsterlaw & Wellman, 2006; Flynn, 2006) and provides us with the opportunity to pinpoint the exact moment of onset of a certain skill. Finally, it would be interesting to explicitly study the relation between the ability to understand about emotions and beliefs and the ability to apply these skills. Regarding future research, it would be interesting to investigate the relation between ToM abilities and cognitive skills such as executive functioning and language over time. Although some developments have been made in this area (Astington & Jenkins, 1999; Carlson et al., 2004; Flynn, 2007), much research has focused on young children only. Investigating the ongoing relationship in older children employing a longer longitudinal study would help to expand our understanding of these relations.

In summary, the results of this study replicate, extend and integrate earlier findings on the development of ToM abilities. In the age range from 5 to 7, children display a significant improvement in both emotion understanding and false-belief understanding. The differences in performance seem to be relatively stable and several interrelations are visible. On the whole, the results seem to suggest that ToM consists of separate but connected skills. These skills show similar developmental trajectories, although some develop at a later age than others. Wellman and Liu (2004) mention Flavell's concepts (1972) of modification or mediation to explain the interconnected developmental trajectories of ToM skills. Rather than simple addition or substitution of skills, earlier skills represent initial insights that broaden over time to later insights. The fact that our tasks did not show complete interrelations also suggests that ToM should be treated as consisting of multiple skills, and that measuring ToM skills by a single concept does not paint a representative picture of the complete ToM skills of an individual.

In addition to the predictive relations within the realm of ToM skills, we found linguistic ability to function as an important predictor. Since all ToM tasks were assessed using verbal tasks, linguistic ability would function to some extent as a prerequisite for successful performance. Lacking further information, it could be argued that the interrelationships between ToM skills within and even between assessment times could all be explained by a common dependence on language skills. However, even after controlling for language skills, these relations remained. This suggests that although language is a prerequisite for ToM performance, it is not enough to account for the variability of ToM performance found in our population and that there are more factors at play.

Since ToM abilities are considered essential for understanding and applying socially appropriate behaviour (Astington & Jenkins, 1995; Beer, Heerey, Keltner, Scabini, & Knight, 2003), timely assessment of ToM problems may help to prevent consequent social repercussions. The stability of performance found in the present study, combined with the measured predictive relations between the abilities over time, provide an indication that measured ToM performance at age 5 is not just a snapshot but can provide information on the long-term outlook as well.

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.

Arbuckle, J. L. (2005). *AMOS 6.0* [computer software]. Spring House, PA: Amos Development Corporation.

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition and Emotion*, *9*, 151–165.

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

Banerjee, M. (1997). Hidden emotions: Preschoolers' knowledge of appearance–reality and emotion display rules. *Social Cognition*, *15*, 107–132.

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme*, *34*, 174–183.

Bartsch, K. (1998). False belief prediction and explanation: Which develops first and why it matters. *International Journal of Behavioral Development*, *22*, 423–428.

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, *85*, 594–604.

Bloom, M. (2003). Theory of Mind and emotion. *Perspectives in Psychology*, *6*, 1–8.

Brown, J. R., & Dunn, J. (1996). Continuities in emotion understanding from three to six years. *Child Development*, *67*, 789–802.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Buitelaar, J. K., Van der Wees, M., Swaab-Barneeld, M., & Van der Gaag, R. J. (1999). Verbal memory and performance IQ predict theory of mind and emotion recognition ability in children with autistic spectrum disorders and in psychiatric control children. *Journal of Child Psychology and Psychiatry*, *40*, 869–881.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*, 1105–1122.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development*, *70*, 853–865.

De Rosnay, M., Pons, F., Harris, F., & Morrell, J. M. B. (2004). A lag between understanding false belief and emotion attribution in young children: Relationships with linguistic ability and mothers' mental-state language. *British Journal of Developmental Psychology*, *22*, 197–218.

Flavell, J. H. (1972). An analysis of cognitive-developmental sequences. *Genetic Psychology Monographs*, *86*, 279–350.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1990). Young children's understanding of fact beliefs versus value beliefs. *Child Development*, *61*, 915–928.

Flynn, E. (2006). A microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology*, *24*, 631–654.

Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development*, *16*, 53–69.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

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

Hughes, C., Adlam, A., Happé, F. G. E., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test–retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry*, *41*, 483–490.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*, 1026–1037.

Kestenbaum, R., & Gelman, S. A. (1995). Preschool children's identification and understanding of mixed emotions. *Cognitive Development*, *10*, 443–458.

Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology*, *20*, 163–190.

Moses, L. J., & Flavell, J. H. (1990). Inferring false beliefs from actions and reactions. *Child Development*, *61*, 929–945.

Mutter, B., Alcorn, M. B., & Welsh, M. (2006). Theory of mind and executive function: Working-memory capacity and inhibitory control as predictors of false-belief task performance. *Perceptual and Motor Skills*, *102*, 819–835.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *TRENDS in Cognitive Sciences*, *3*, 337–344.

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.

Pons, F., & Harris, F. (2005). Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. *Cognition and Emotion*, *19*, 1158–1174.

Repacholi, B., & Slaughter, V. (Eds.). (2003). *Individual differences in theory of mind: Implications of typical and atypical development*. New York: Psychology Press.

Saarni, C. (1979). Children's understanding of display rules for expressive behavior. *Developmental Psychology*, *15*, 424–429.

Underwood, M. K., Coie, J. D., & Herbsman, C. R. (1992). Display rules for anger and aggression in school-age children. *Child Development*, *63*, 366–380.

Verhoeven, L., & Vermeer, A. (2001). *Taaltest Alle Kinderen (TAK)* [Language Test for all Children]. Arnhem, The Netherlands: CITO.

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

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541.