CHILD DEVELOPMENT

Child Development, September/October 2011, Volume 82, Number 5, Pages 1691–1703

Developmental Continuity in Theory of Mind: Speed and Accuracy of Belief–Desire Reasoning in Children and Adults

Ian A. Apperly
University of Birmingham, UK

Frances Warren
Reading University, UK

Benjamin J. Andrews, Jay Grant, and Sophie Todd
University of Birmingham, UK

On belief–desire reasoning tasks, children first pass tasks involving true belief before those involving false belief, and tasks involving positive desire before those involving negative desire. The current study examined belief–desire reasoning in participants old enough to pass all such tasks. Eighty-three 6- to 11-year-olds and 20 adult participants completed simple, computer-based tests of belief–desire reasoning, which recorded response times as well as error rates. Both measures suggested that, like young children, older children and adults find it more difficult to reason about false belief and negative desires than true beliefs and positive desires. It is argued that this developmental continuity is most consistent with either executive competence or executive performance accounts of the development of belief–desire reasoning.

Much research in the last 30 years has examined when and whether young children and nonhuman animals have a ''theory of mind'' (ToM) (e.g., Call & Tomasello, 2008; Doherty, 2008; Premack & Woodruff, 1978; Wimmer & Perner, 1983). The same ability to reason about mental states such as beliefs, desires and intentions has also recently begun to receive attention from cognitive neuroscientists and cognitive psychologists, whose main experimental participants have been human adults (e.g., Apperly, Samson, & Humphreys, 2009; Frith & Frith, 2003; Saxe, Carey, & Kanwisher, 2004).

Between these two literatures there remains a significant gap in our understanding of basic ToM processes in ''older'' children, over the age of 5 or 6 years. This gap exists because ToM in children has traditionally been studied by testing children's accuracy on simple reasoning tasks, which are typically passed by 4 or 5 years of age. Although some research has examined the abilities of older children on more complex or subtle ToM tasks (e.g., Baron-Cohen, O'Riordan, Stone, Jones, & Plaistead, 1999; Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001; Happe, 1994), the divergence between these methods and those used with younger children makes it difficult to test for continuities or discontinuities in development.

What is needed are ways to test participants who ''pass'' ToM tasks on the same, simple ToM processes examined in young children. This problem has recently been solved in several studies of adult participants that use response time (RT), as well as accuracy, as their dependent variables (for a review, see Apperly et al., 2009). The current study adopts the same approach in order to study simple belief–desire reasoning in 6- to 11-year-old children, who would be expected to ''pass'' simple belief–desire reasoning tasks.

Belief–Desire Reasoning

Most research on ToM has concerned how and when children first attribute beliefs and desires successfully. There is evidence that 3-year-olds are able to reason that people with different beliefs (e.g., about the location of a puppy) will behave in different ways, provided the child does not know

The research reported here was conducted with support from a research bursary from the Experimental Psychology Society, UK.

Correspondence concerning this article should be addressed to Ian Apperly, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. Electronic mail may be sent to i.a.apperly@bham.ac.uk.

 0009-3920/2011/8205-0026
DOI: 10.1111/j.1467-8624.2011.01635.x

whether these beliefs are true or false (e.g., Wellman & Bartsch, 1988; Wellman & Liu, 2004). However, it is well established that false belief tasks are significantly harder (e.g., Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004). In one such task participants witness ''Maxi'' putting his chocolate in a cupboard and then leaving it unattended (Wimmer & Perner, 1983). In his absence, his mother moves the chocolate to a drawer, resulting in Maxi holding a false belief about its location. Consequently, when asked where Maxi will first look for his chocolate, most 4- and 5-year-olds judge correctly that he will return to the chocolate's original location, where he mistakenly believes his chocolate to be. However, younger children tend to judge that Maxi will know to look for the chocolate in its current location. These findings have led to suggestions that children understand true beliefs, or ''diverse'' beliefs (whose truth is unknown) before they understand false beliefs (e.g., Wellman & Liu, 2004).

As young as 18 months there is evidence that children understand that people may have different preferences (e.g., Repacholi & Gopnik, 1997), and by 3 years children can predict actions on the basis of what an agent desires, even when this conflicts with the child's own desire (e.g., Rakoczy, Warneken, & Tomasello, 2007; though see, e.g., Moore et al., 1995; Perner, Zauner, & Sprung, 2005, for opposing views). Interestingly, however, there is evidence of inter-actions between the demands of reasoning about beliefs and desires. For example both Cassidy (1998) and Leslie and Polizzi (1998) found that children who passed a false belief task when the character's desire was to obtain the hidden object were prone to fail when the character's desire was to avoid the object, with success on the latter task not occurring reliably until 6 years or older (e.g., Leslie, Friedman, & German, 2004; Friedman & Leslie, 2005).

In sum, the literature suggests that children consistently pass tasks involving true beliefs (B+) before tasks involving false belief (B−), and tasks involving a positive desire for an object (D+) before tasks involving a negative desire (D−). Performance on such tasks is also reliably associated with children's performance on tests of ''executive function,'' which assess their ability to behave in a flexible, goal-oriented manner (e.g., Carlson & Moses, 2001; Carlson, Moses, & Hix, 1998; Perner & Lang, 1999; Sabbagh, Moses, & Shiverick, 2006), and can be improved somewhat by task variations that attempt to reduce executive demands or clarify confusing wording (e.g., Mitchell & Lacohee, 1991; Siegal & Beattie, 1991; Yazdi, German, Defeyter, & Siegal, 2006). However, it is unclear what happens once children pass these tasks: Do the differences in difficulty persist or do they largely disappear as children gain practice? This question is important for two reasons. First, testing whether belief–desire reasoning continues to pose challenges for older children can provide evidence for or against alternative accounts of why belief–desire reasoning problems are of varying difficulty for younger children. Second, understanding any challenges that belief–desire reasoning continues to pose for older children might shed light on continuing development in ToM after children pass the most developmentally sensitive tasks.

Why Do Belief–Desire Reasoning Problems Vary in Difficulty for 3- to 5-Year-Old Children?

There are several accounts of the changes in children's belief–desire reasoning between 3 and 5 years that differ in their capacity to speak to questions about the relative difficulty of belief–desire reasoning problems in older children and adults.

A common interpretation of changes in 3- to 5-year-old children's performance on belief–desire reasoning tasks is that children's conception of mental representation undergoes a fundamental change during the preschool years, (Gopnik & Wellman, 1992; Perner, 1991; Wellman et al., 2001). However, there are very different accounts of the role of executive function in these conceptual changes. ''Pure conceptual change'' accounts (e.g., Gopnik & Wellman, 1992; Perner, 1991; Wellman et al., 2001) afford executive function (or other processing factors) a role in managing the incidental demands of ToM tasks (such as remembering the sequence of events in the story or preventing the child from giving the same response as on an earlier question) but changes in executive capacity do not form part of the explanation for why children pass some tasks (e.g., B+D+) earlier than others (e.g., B−D+). In contrast, ''Executive emergence'' accounts (e.g., Carlson & Moses, 2001; Moses, 2001) hold that executive function enables the emergence of ToM concepts. This casts executive function in the role of scaffolding in the construction of a building, insofar as executive function is necessary for the process of construction but does not form part of the finished product (Apperly et al., 2009). Such accounts might explain the incremental progress in children's passing of belief–desire tasks by supposing that incremental increases in executive capacity enabled the acquisition of increasingly complex ToM concepts. Importantly, however, neither pure conceptual accounts nor executive emergence accounts would predict continuing differences in difficulty on different types of belief–desire task: Once the necessary concepts are firmly in place, there is no reason to believe that later acquired concepts should be harder to deploy than earlier acquired concepts.

Another way of conceptualizing the relation between executive function and ToM is to suppose that a certain level of executive function is necessary in order for anyone—young children, older children, or adults—to have conceptual thoughts about mental states. (e.g., Russell, 1996, 1999). Continuing the building metaphor, these accounts place executive function in the role of mortar, rather than scaffolding, in that it is necessary for construction *because* it is an intrinsic part of the finished product. Like emergence accounts, such ''executive competence'' accounts explain the order of conceptual development in terms of varying executive demands for concept acquisition. Unlike emergence accounts, however, mental state concepts do not float free from their dependence on executive function once they have been acquired. Consequently, executive competence accounts clearly predict that later acquired concepts (such as false belief) should continue to be harder to entertain than earlier acquired concepts (such as true belief) in older children and adults.

Finally, *executive performance* accounts argue that executive function is critically involved in the deployment of preexisting mental state concepts. (e.g., Carlson & Moses, 2001; Friedman & Leslie, 2004, 2005; Leslie, German, & Polizzi, 2005; Leslie & Polizzi, 1998; Mitchell, 1996; Russell, Mauthner, Sharpe, & Tidswell, 1991). The most specific claims of this kind have been made by Leslie and colleagues, who argue that both false belief and negative desire make demands on children's inhibitory control, with the greatest demands when these mental states occur in combination (i.e., in B−D− tasks). Although the specific role for executive function is rather different, in terms of the building metaphor, executive performance accounts resemble executive competence accounts, insofar as they place executive function in the role of mortar rather than scaffolding. Because of this, like the executive competence account, the executive performance account of Leslie and colleagues clearly predicts that the need for inhibition (and thus the gradation of difficulty between B+D+ tasks and B−D− tasks) should persist in older children and adults.

Continuing Development in ToM

Although most studies have investigated ToM in 3- to 5-year-olds, few people who have actually met a 5-year-old would doubt that they fall short of having a mature ability to reason about the minds of others. There is some evidence of continuing developments in the sophistication of children's ToM abilities (e.g., Baron-Cohen et al., 1999; Chandler, Boyes, & Ball, 1990; Happe, 1994; Perner & Wimmer, 1985) and in the ability to resist egocentric interference during communication (Dumontheil, Apperly, & Blakemore, 2010; Epley, Morewedge, & Keysar, 2004). There are also reasons for supposing that even the more basic ToM abilities investigated in young children may continue to change, because of late maturation of the neural systems implicated in simple ToM abilities (Blakemore, 2008; Blakemore & Choudhury, 2006; Giedd et al., 1999; Paus, Evansm, & Rapoport, 1999; Reiss, Abrams, Singer, Ross, & Denckla, 1996), and because of continuing developments in executive function and processing speed (e.g., Luna, Garver, Urban, Lazar, & Sweeney, 2004). However, to date there has been little direct investigation of simple ToM processes in older children.

Evidence from adults can inform accounts of children's developing ToM by providing information about the terminus of development, and such studies suggest several roles for executive processes in fully developed ToM abilities (e.g., Apperly et al., 2009). A study by German and Hehman (2006) is of particular relevance in the current context. In this study adult participants were presented with short stories. A subsequent question about a character in the story required participants to infer the character's belief (true or false), and integrate this with his/her desire (to obtain or avoid an outcome) in order to predict their action. Adults were both slower and more error prone when faced with combinations of belief and desire that children find more difficult (e.g., B−D−) compared to ones that young children pass at a younger age (e.g., B+D+). Moreover, individual differences in adults' performance were correlated with independent measures of processing speed and inhibitory control.

These and other findings suggest that later developing ToM abilities are more difficult for adults than earlier developing abilities and suggest that executive function continues to have a role in ToM in adults. However, it remains important to gain direct evidence from children over the age of 5 rather than just from adults. One reason is that it is not entirely clear that errors in 3- to 5-year-olds and

RT differences in adults reflect the same underlying processes, because the tasks and the measures used are rather different. Much stronger evidence for developmental continuity would come from assessment of the error rates and RTs of older children and adults assessed on very similar tasks. Another reason why it is important to gain evidence from older children is that evidence from the RTs of adults tells us little about how the processing of ToM problems may have changed since children first passed ToM tasks. Does later age-related change consist only in overall improvements in speed and accuracy, or also in a reduction in the relative difficulty of different ToM problems?

The Current Study

The first aim of the current study was to bridge the gap between existing developmental studies of 3- to 5-year-olds' belief–desire reasoning and German and Hehman's (2006) study of adults. To do this we tested children aged between 6 and 11 years, as well as an adult sample, on a task that involved all four permutations of belief–desire reasoning described above: B+D+, B+D−, B−D+, B−D−. Most of the children and all of the adults were expected to have the concepts of belief and desire necessary for them to pass the belief–desire reasoning tasks used with 3- to 5-year-olds, up to and including B−D− tasks. Our interest was therefore primarily in RTs to the four conditions, as well as any residual errors.

The second aim of the current study was to test performance on a belief–desire reasoning task that did not require a belief inference. The vast majority of research on ToM in children and adults requires participants to infer the mental states (such as the false belief) of a character from observation of their behavior, or from the unfolding narrative. However, there is evidence that even when preschoolers and adults are simply told about a character's false beliefs (so they need make no inference themselves) they show a tendency for egocentric interference that resembles the difficulties observed on more standard false belief tasks (e.g., Apperly, Back, Samson, & France, 2008; Wellman & Bartsch, 1988). This is of some theoretical importance because it is often supposed that the process of inferring a belief is one of the key processing steps that makes demands on inhibitory control (e.g., Leslie et al., 2005). In the current study we obviated the need for any belief inference to be made by simply telling participants where a character thought an object was located, where it was really located and whether or not the character liked the object. Finding out whether a difference remains between true and false belief conditions in these circumstances helps to narrow in upon the processes in belief–desire reasoning that are likely sources of difficulty for children.

Method

For both children and adults, each trial of the experiment required participants to judge which of two boxes a male cartoon character would open. They were told that one of the two boxes contained some food. Sometimes the character liked the food and sometimes he disliked the food. If he liked the food he would open the box that he thought contained the food, whereas if he disliked the food he would try to open the other box. Participants were presented with information about which box contained the food (''reality''), which box the character thought contained the food (''belief''), and whether or not he liked the food (''desire''). The dependent variables were the speed and accuracy of participants' deduction about which box the character would open.

Participants

Children. Twenty-five 6- to 7-year-olds (14 female, 11 male; mean age = 6;10, range = 6;9 to 7;2), thirty-two 8- to 9-year-olds (13 female; mean age = 8;8, range = 8;2 to 9;2), and twenty-six 10- to 11-year-olds (14 female; mean age = 11;0, range = 10;6 to 11;4) participated. Children were sampled randomly from three classes, each two school years apart. All children spoke English as their first language and were predominantly White Caucasian (95%), with a small number from an Asian background, and were from a middle-class background.

Adults. Twenty undergraduate adult participants (16 female, 4 male; mean age = 21 years, range = 18–35) participated for course credits or a small honorarium. All spoke English as their first language and were predominantly White Caucasian (80%), with a small number from Asian and African backgrounds.

Materials

Children and adults were tested on standard laptop and desktop computers respectively, using E-Prime software to present the stimuli

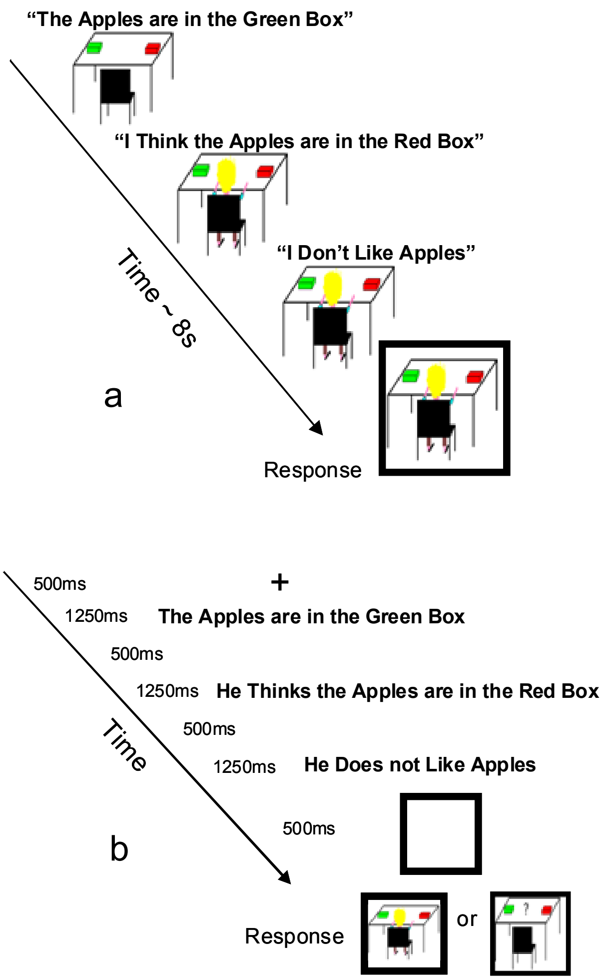

Figure 1. Schematic event sequences from a trial of the method used with children (a) and the method used with adults (b). *Note.* In both cases the depicted sequence is a B–D– trial.

(http://www.pstnet.com/products/eprime). Sixteen different foods (carrots, grapes, biscuits, apples, cakes, raisins, mushrooms, peas, cherries, peanuts crackers, crisps, muffins, cookies, peaches, and yogurt) were used throughout the trials.

Design and Procedure

Children. Each trial was composed of three sentences played as sound files, accompanied by three pictures (see Figure 1a). For the first sentence the male character was absent from the picture and a female voice described the real location of the food item (e.g., ''The apples are in the green box''). This sentence was always presented first as pilot work suggested that children found this the easiest to understand, regardless of experimental condition. For the second and third sentences, the male character was present in the picture and a male voice, speaking in the first person, reported the character's belief about the food's location and his food preference (e.g., ''I think the apples are in the green/red box''; ''I like/don't like apples''). The duration of each sentence ranged from 1500 to 2500 ms, and each sentence was separated from the next by 500 ms of silence. After sentence 3, a black square appeared around the picture cueing children to predict which box the character would open.

Half of the time the male character had a true belief (his belief statement corresponded to the reality described in sentence 1) and half of the time he had a FB. Half of the time the man expressed a liking for the food and half of the time he said he did not like the food. Half of the time the man's belief was expressed in sentence 2 and half of the time in sentence 3 (with the corresponding opposite pattern for the sentence describing his like or dislike of the food). The factors of belief and desire (liking or disliking) were crossed to generate four experimental conditions, B+D+, D+D−, B−D+ and B−D−, with four items in each condition, leading to a total of 16 experimental trials. These experimental trials were presented in two equivalent blocks containing two trials of each condition. Each child was presented with trials in one of four pseudorandom orders that avoided more than two consecutive trials of the same type with the same sentence order, or with the same position of green and red boxes.

Children were tested individually in a quiet room. The experimenter talked children through a set of instructions with accompanying pictures, explaining that the aim of the task was to predict which box the character will open. Children completed a practice trial from each of the four experimental conditions and received corrective feedback. Children who made errors at this point were encouraged to repeat the practice session. Only one participant made an error on the second round of the practice session. Although continuing with the test trials, his scores were excluded from the data set.

For the experimental blocks participants were asked to start each trial with both index fingers located on a central point (15.5 cm from both highlighted response buttons). Once the experimenter was satisfied that the child was concentrating they initiated the trial by pressing the keyboard. Subsequently the trial progressed automatically through three sentences and ended when the child pressed the key corresponding to the box that they predicted the character to open. The key-press and the RT were recorded automatically by the computer.

Adults. The method used with children was adapted for use with adults. Trials began with a fixation cross for 500 ms. The three sentences of each trial (e.g., ''He thinks the apples are in the green box''; ''He does like apples''; ''The apples are in the red box'') were presented sequentially as text, for 1250 ms each, and without accompanying pictures (see Figure 1b). There followed a blank screen for 1250 ms, then an empty black frame appeared for 500 ms, acting as a warning that the test stimulus was about to appear. On two thirds of the trials the test stimulus was one of the two pictures from the method used with children, depicting the cartoon man facing two boxes on the table, with the red box either on the left or the right. Note that this was the first point at which adults knew the locations of the red and green boxes and so was the first point at which they could plan their key-press. Pilot work suggested that this change from the procedure used with children was necessary because adults tended to plan key-presses in advance, whereas children did not. On the other one third of trials, the response stimulus depicted an empty table with a question mark positioned between the two boxes. This signaled that participants should indicate the location of the object rather than the box in which the man would search. Once again, pilot work indicated that without these catch trials, adults (but not children) exploited a strategy of ignoring reality entirely.

Pilot work indicated that adults had no difficulty understanding the three sentences presented in any order, and so belief, desire, and reality sentences were presented equally often in all possible orders. Otherwise, stimuli were subject to the same process of counterbalancing as for the method used with children. The key difference was that adults were presented with 296 trials in four equivalent blocks. The same 16 food items were used repeatedly, but the man's like or dislike of a particular food was fixed.

Adult participants were tested in the laboratory. The experimenter explained the task, and the participant completed a block of six practice trials before entering the main experiment.

Results

Children's Data

Response times. All RTs were recorded from the earliest point in sentence 3 at which a correct response could be initiated. When sentence 3 described beliefs (e.g., ''I think the mushrooms are in the red box'') RTs were measured from the beginning of the word *box*. When sentence 3 described desire (e.g., ''I don't like mushrooms'') RTs were measured from the beginning of the word *like*. (Similar results were obtained when we measured RTs from the beginning of the word *mushrooms*.) All incorrect responses were removed from the data set. This resulted in a loss of some data points: 97 (24.3%) from the 6- to 7-year-olds; 67 (13.5%) from the 8- to 9-year-olds; and 45 (11%) from the 10- to 11-year-olds. Of the remaining data, any RT falling more than 3 *SD* outside of the condition mean was similarly rejected, resulting in the removal of further data points: 15 from the 6- to 7-year-olds; 12 from the 8- to 9-year-olds; 9 from the 10- to 11-year-olds. Finally, if a participant had no data points for a condition, their remaining data points in other conditions were not included in the analysis. This resulted in the loss of 13 participants from the 6- to 7-year-old group, 10 participants from the 8- to 9-year-old group and 4 participants from the 10- to 11-year-old group. (We conducted an alternative analysis in which all children were retained and missing data for each condition were interpolated from the condition means. This analysis revealed the same pattern of significant and nonsignificant results as the analysis presented below.)

A split measures analysis of variance (ANOVA) was conducted on the remaining data, using age as the between-subjects factor (6–7 years, 8–9 years, and 10–11 years), and belief (B+, B−), desire (D+, D−), and sentence order (last sentence heard = desire or belief) as the within-subjects factors. This revealed significant main effects of belief, $F(1, 52) = 21.7$, $p < .001$, $\eta^2 = .294$: B+ < B−; desire, $F(1, 52) = 61.45$, $p < .001$, $\eta^2 = .542$: D+ < D−; order, $F(1, 52) = 24.39$, $p < .001$, $\eta^2 = .319$ (quicker responses when belief sentences were presented last); and age, $F(1, 52) = 17.59$, $p < .001$, $\eta^2 = .403$ (10- to 11-year-olds < 8- to 9-year-olds < 6- to 7-year-olds). Mean RTs for each condition and each age group are charted in Figure 2.

Significant interactions were further obtained between desire and age, $F(1, 52) = 4.057$, $p = .023$, $\eta^2 = .135$, and between order and desire, $F(1, 52) = 4.22$ $p = .045$, $\eta^2 = .075$. All other interactions were nonsignificant: largest nonsignificant, $F(1, 52) = 1.67$, $p = .198$, $\eta^2 = .06$. The significant interaction between desire and age suggests that the effect of the two desire combinations differ across the age groups. To investigate this interaction further, the two desire combinations (D+ and D−) were collapsed across the different orders and entered into paired-sample t tests. All comparisons reached

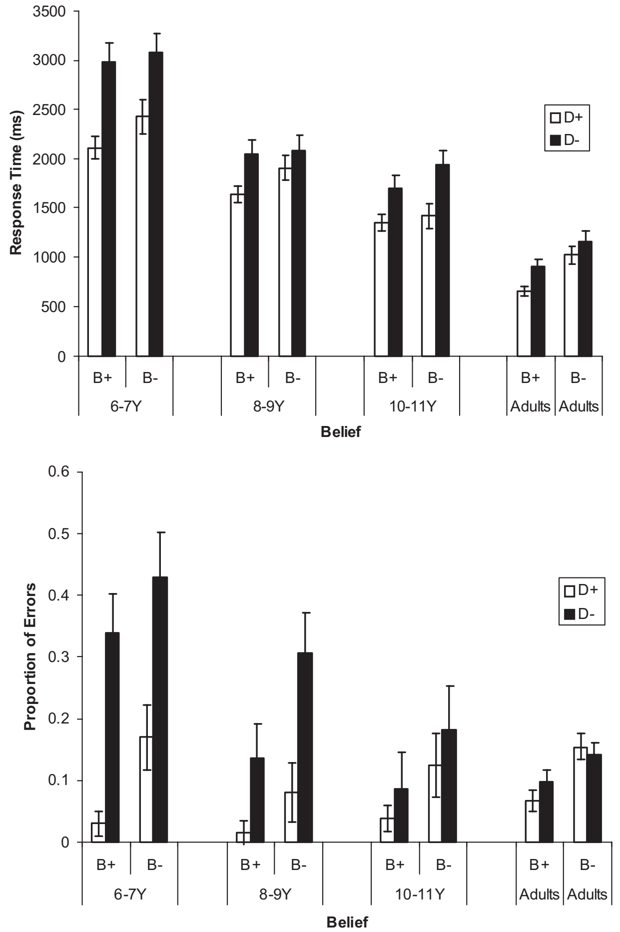

Figure 2. Mean response times (top panel) and proportion of errors (bottom panel) for the four types of belief–desire reasoning trial. Error bars denote ±1 SE.

significance, indicating that at all ages, D− conditions resulted in reliably slower RTs in comparison to D+ conditions: 6- to 7-year-olds, $t(11) = 3.36$, $p = .006$; 8- to 9-year-olds, $t(20) = 4.78$, $p < .001$; and 10- to 11-year-olds, $t(21) = 6.17$, $p < .001$. From Figure 2 it is apparent that the 8- to 9-year-old group shows the smallest difference in RTs between D+ and D− trials.

To investigate the interaction between order and desire, paired-samples *t* tests were conducted to compare RTs on D+ and D− trials in each order. There was a significant effect of desire for both orders: desire sentence last, $t(53) = 7.78$, $p < .001$, and belief sentence last, $t(53) = 3.41$, $p = .001$, but this effect was larger when the desire sentence was heard last (D+ = 1792 ms, D− = 2345 ms) than when the belief sentence was heard last (D+ = 1667 ms, D− = 1993 ms).

Error rate. To study the effects of the four belief–desire combinations on error rate, the mean proportion of incorrect scores was calculated for each condition across the two different orders. To check that children were able to perform the task we first assessed whether each age group's performance was significantly above chance for each of the four belief–desire conditions, irrespective of order. One-sample *t* tests were used to compare each condition against a theoretical chance baseline of 50% correct. The 6- to 7-year-old group was not above chance for the B−D− condition, $t(24) = 1.0$, $p = .327$. All groups were above chance for all other conditions (all *t*s > 2.18, all *p*s < .039).

For the main analysis, error rates were entered into a split-measures ANOVA with age as the between-subjects factor, and belief, desire and order as the within-subjects factors. This revealed significant main effects of belief, $F(1, 79) = 19.03$, $p < .001$, $\eta^2 = .194$: B+ < B−; desire, $F(1, 79) = 23.93$, $p < .001$, $\eta^2 = .233$: D+ < D−; and age, $F(1, 79) = 4.41$, $p = .015$, η^2: 10- to 11-year-olds < 8- to 9-year-olds < 6- to 7-year-olds, but not of order, $F(1, 79) = 2.09$, $p = .153$), $\eta^2 = .026$. Mirroring the effects demonstrated with the RT data, and as indicated in Figure 2, these significant findings suggest that conditions involving either a false belief or a negative desire elicited more errors. This evidence, converging with that from the RT data, suggests that the overall results found were not simply the consequence of a trade-off between speed and accuracy; the manipulations associated with the highest error rate were those on which participants responded slowest.

A significant interaction was also found between age and desire, $F(1, 79) = 3.49$, $p = .035$, $\eta^2 = .081$, but all other interactions were nonsignificant: largest nonsignificant, $F(1, 79) = 1.99$, $p = .143$, $\eta^2 = .048$. To investigate the interaction between desire and age we compared error rates in the D+ and D− conditions for each age group. Paired-samples *t* tests were significant for both the 6- to 7-year-old data, $t(24) = -3.863$, $p = .001$, and the 8- to 9-year-old data, $t(30) = -3.090$, $p = .0040$, but not for the 10- to 11-year-old data, $t(25) = 1.066$, $p = .297$. While the younger two age groups were significantly more error prone in the D− conditions, this was not the case for the oldest participant group.

Adult Data

Response time. As with the children's data, all incorrect responses were removed from the data set, resulting in a total loss of 442 data points (11.5%). Of the remaining data, RTs falling 3 *SD* away from the condition mean were also rejected, leading to a further loss of 58 data points. All adult participants had remaining data points for each condition, so the all participants were entered into the main analysis. A repeated measures ANOVA was conducted, with belief, desire, and order (last sentence read = belief, desire, or reality) as within-subject factors. This analysis revealed significant main effects of belief, $F(1, 19) = 24.65$, $p < .001$, $\eta^2 = .565$: B+ < B−, and desire, $F(1, 19) = 25.88$, $p < .001$, $\eta^2 = .577$: D+ < D−, but not of order, $F(1, 19) = 0.614$, $p = .547$, $\eta^2 = .031$. In addition, a significant interaction was obtained between belief and desire, $F(1, 19) = 4.658$, $p = .044$, $\eta^2 = .197$. All other interactions were nonsignificant: largest nonsignificant, $F(1, 19) = 2.24$, $p = .121$, $\eta^2 = .105$.

The significant interaction between belief and desire was investigated with paired-samples *t* tests. There was a significant effect of belief (B+ vs. B−), at both levels of desire, both $ts(19) > 3.15$, both $ps < .005$, and a significant effect of desire (D+ vs. D−) when beliefs were true, $t(19) = 9.24$, $p < .001$. However, when beliefs were false, the effect of desire was smaller, $t(19) = 2.13$, $p < .047$, and this was nonsignificant if a correction was made for four post hoc comparisons.

Error rate. The effects of the belief and desire on error rate were analyzed in the same way as the children's data. The mean proportion of incorrect scores was again calculated for each condition across the three different orders, and these data showed a similar overall level of errors to that reported by German and Hehman (2006). These were entered into an ANOVA, using the same within-subjects factors as used in the RT analysis. There was a significant main effect of belief, $F(1, 19) = 17.62$, $p < .001$, $\eta^2 = .481$: B+ < B−, but not of desire, $F(1, 19) = 0.281$, $p = .602$, $\eta^2 = .015$, or order, $F(1, 19) = 0.739$, $p = .484$, $\eta^2 = .037$. Significant interactions were found between belief and order, $F(1, 19) = 13.02$, $p < .001$, $\eta^2 = .407$, as well as between belief, desire, and order, $F(1, 19) = 3.37$, $p = .045$, $\eta^2 = .151$. The other interactions were nonsignificant.

To investigate the three-way interaction between belief, desire, and order, the effects of belief and desire were analyzed separately for each order in 3 two-way repeated measures ANOVAs. When reality was the last sentence presented, there were no main effects of belief, $F(1, 19) = 0.733$, $p = .403$, $\eta^2 = .037$, or desire, $F(1, 19) = 0.986$, $p = .333$, $\eta^2 = .049$, and there was no significant interaction, $F(1, 19) = 0.616$, $p = .442$, $\eta^2 = .031$. When desire was the last sentence presented there was a significant main effect of belief, $F(1, 19) = 31.020$, $p < .001$, $\eta^2 = .620$; B+ < B−, but not for desire, $F(1, 19) = 0.725$, $p = .405$, $\eta^2 = .037$. The interaction between the two factors was also nonsignificant, $F(1, 19) = 0.228$, $p = .639$, $\eta^2 = .012$. With belief as the last sentence, a main effect of belief was revealed, $F(1, 19) = 7.11$, $p = .015$, $\eta^2 = .272$; B+ < B−, but not of desire, $F(1, 19) = 0.406$, $p = .531$, $\eta^2 = .021$. There was also a significant interaction between belief and desire, $F(1, 19) = 7.64$, $p = .012$, $\eta^2 = .287$. In order to examine this interaction further, the individual condition means from within this order were compared. The error rate in B−D+ trials (0.19/1) was significantly higher than in either B+D+ trials (0.07/1), $t(19) = -3.44$, $p = .003$, or B−D− trials (0.11/1), $t(19) = -2.56$, $p = .019$, whereas the error rate in B+D− trials (0.12/1) did not differ from the error rate in B−D− trials, $t(19) = .348$, $p = .732$, or B+D+ trials, $t(19) = 1.20$, $p = .244$.

Discussion

This study is the first to chart changes in children's basic belief–desire reasoning after they have passed the suite of tasks most commonly used with 3- to 5-year-old children. There were several notable findings. First, belief–desire reasoning tasks that are harder for 3- to 5-year-olds (those involving B− and D−) continue to be harder for older children, and, consistent with German and Hehman (2006), this pattern also holds in adults. Second, this pattern is observed even when participants do not need to infer a character's belief but must simply deduce the character's action from given information about his belief and his desire. Third, there were large improvements in speed and accuracy with increasing age. We discuss each of these findings in turn.

Alternative Accounts of the Development of Belief–Desire Reasoning

Existing studies of 3- to 5-year-olds suggest that different kinds of belief–desire reasoning tasks are first passed at different ages, beginning with cases where the agent has a true belief and a positive desire (B+D+) and ending with cases where the agent has a false belief and a negative desire (B−D−). In the current study we began testing the speed and accuracy of belief–desire reasoning in children who were just above the usual age of interest in research on ToM, and we used a method that could be employed with older children and, with some adaptation to counter strategies, with adults.

The youngest children in our sample, aged 6–7 years, still made a relatively large number of errors on our task. Two factors may have contributed to this. First, although the literature suggests that most children pass most belief–desire reasoning tasks by 6–7 years, it is uncommon for an entire sample to be at ceiling (see, e.g., Wellman et al., 2001), so at least some of the errors observed are likely due to children who do not yet have a firm grasp of all four belief–desire reasoning conditions. Second, and more important, our task was not designed to assess the earliest age at which success would be demonstrated but rather to assess the processing cost involved in the four belief–desire reasoning conditions. It is notable that even adults' accuracy was not at ceiling. This is normal in studies of adults (see, e.g., Apperly et al., 2009, for a recent review) yet we do not doubt that adults had a firm grasp of all four belief–desire reasoning conditions. Rather, the errors of adults, as well as most of the errors of children in our samples, should be interpreted together with RTs as an index of the overall processing cost of the different experimental conditions.

In light of this, it is critical that the error pattern we observed was not random: This indicates that children and adults were not simply confused or overwhelmed by the general demands of interpreting the stimulus sentences, remembering the location of the hidden object that was never actually visible to participants, and deducing the action of the character. Instead, at all ages the error rates and RTs showed the same patterns of difficulty observed when younger children first pass these tasks. In our youngest children there were more errors on trials involving B− than those involving B+, and more errors on trials involving D− than those involving D+. Where we had sufficient data, analysis of these children's RTs revealed the same pattern of relative difficulty. With increasing age participants made fewer errors and responded more quickly, but the same pattern of difficulty (B+ < B−, D+ < D−) held throughout.

Of course, we must be cautious about drawing direct comparisons between the patterns observed in the current study and previous studies with younger children. It is clear that the current methods were rather different from those used with younger children, and it cannot be taken for granted that they are assessing the same underlying abilities. However, in terms of the requirements on belief–desire reasoning, the judgments required in our task are logically identical to those required in tasks used with younger children. This provides a theoretical basis for similarity that fits well with the empirical observation that the pattern of difficulty across conditions in our study is in fact similar to the patterns observed with younger children using different methods. With the important caveat that this still falls short of a direct demonstration of continuity, we go on to discuss our findings on the assumption that our task does allow legitimate comparisons to be made between our findings and those from studies of younger children.

With the above caveat, our findings hold implications for existing theories of the development of ToM in 3- to 5-year-olds. As described in the Introduction, neither pure conceptual change accounts (e.g., Bartsch & Wellman, 1995; Gopnik & Wellman, 1992; Perner, 1991) nor executive emergence accounts (e.g., Carlson & Moses, 2001) predict continuing variation in the difficulty of different kinds of belief–desire reasoning once the relevant concepts are firmly in place. As a result, the current finding that the pattern of processing demands in older children and adults exactly matches the pattern in which younger children first succeed on different belief–desire reasoning problems would have to be explained by some additional means. Of course, there is nothing in either the pure conceptual change account or the executive emergence account that rules out such additions. It is possible to posit that some aspects of belief–desire reasoning are later to acquire because they are more conceptually complex, or because their acquisition requires more executive control, *and in addition* that these same aspects of belief–desire reasoning continue to be more difficult for older children and adults because they are also more difficult to represent or to process. However, it is not possible for these accounts to posit that some aspects of belief–desire reasoning are later to acquire than others *because* they are more difficult to represent or to process, or rather, if they do so then they become indistinguishable from executive competence accounts.

In contrast, executive competence and executive performance accounts of 3- to 5-year-olds' ToM development directly predict continuing variation in the difficulty of different belief–desire reasoning problems. Executive competence accounts (e.g., Russell, 1996, 1999) retain the assumption that there is genuine conceptual change in the ToM abilities of 3- to 5-year-olds, but argue that the capacity to entertain these concepts is determined by the child's capacity to deploy working memory and other executive resources in flexible conceptual thought. Some abilities appear later than others (e.g., B− appears later than B+) because they make greater executive demands, and it is natural to expect that they continue to make greater demands in older children and adults. Executive performance accounts contrast with executive competence accounts insofar as they assume that children are innately equipped with concepts of belief and desire, or that by the time children tackle standard laboratory tasks (e.g., FB tasks) they have already acquired the necessary concepts. But critically, like executive competence accounts, they also propose that some ToM abilities are manifest later than others because putting these concepts into action makes higher demands in some circumstances than in others (e.g., Friedman & Leslie, 2004, 2005; Leslie & Polizzi, 1998; Leslie et al., 2005). Thus, on both executive competence and executive performance accounts the similarity in the order in which young children succeed on different belief–desire reasoning tasks and the pattern of processing costs of these tasks in older children and adults does not require any additional explanation; it arises quite naturally out of the claim that different belief–desire reasoning tasks vary intrinsically in the demands they make on executive function.

In sum, the current findings clearly do not rule out pure conceptual change accounts or executive emergence accounts of ToM development. However, if we wish to explain the bigger picture that includes young children's acquisition of ToM abilities, continuing changes through later development, and the end state of development in adults, then these accounts are less parsimonious than either the executive competence or the executive performance accounts.

What Aspects of Belief–Desire Reasoning Are Cognitively Demanding?

Most studies of ToM in young children and in adults use tasks that make multiple demands on participants. For example, standard false belief tasks require young children to infer someone's false belief, to keep this information in mind, and often, to use it to formulate a prediction of what the person will do (e.g., he will search in the incorrect location). With such tasks it is impossible to ask finer grained questions about exactly which processes are cognitively demanding. Nonetheless, some accounts have advanced theoretical ideas about the role of executive function in specific aspects of ToM tasks. For example, Leslie and colleagues have suggested that at least one reason why B+ tasks are easier than B− tasks is that inhibition is needed in the process of *inferring* someone's belief (e.g., Friedman & Leslie, 2004, 2005; Leslie & Polizzi, 1998; Leslie et al., 2005). Because the current task obviated the need for a belief inference (because participants were simply told what the character believed) and because B+ conditions remained easier than B− conditions, the current findings cannot be explained in this way.

The current findings therefore add to a number of other results from children and adults suggesting that ToM tasks may make predictable processing demands even when mental states do not need to be inferred, but must simply be held in mind and used to formulate a response to a test question (e.g., Apperly et al., 2008; de Villiers & Pyers, 2002; Flavell, Flavell, Green, & Moses, 1990; Wellman & Bartsch, 1988). This does not speak directly against any current theory of ToM development, but clearly suggests that more precision is necessary in devising tasks that allow different ToM processes—for example, inferring new mental states versus deducing the consequences of those mental states for an agent's behavior—to be studied independently. It would be informative, both for accounts of early ToM development, and for accounts of ToM in older children and adults, to know whether inferring mental states makes different demands on executive function from holding this information in mind or using it to predict an agent's behavior.

Continuing Development in ToM

Continuing changes in belief–desire reasoning may help explain why older children are better at deploying their basic ToM abilities than younger children even if both older and younger children already pass the most developmentally sensitive belief–desire reasoning tasks. By far the most notable age-related change in the current study was that older children were significantly faster and more accurate at all belief–desire reasoning trials than younger children. This may seem a predictable result, given age-related change in processing speed (e.g., Luna et al., 2004) but the potential significance of these changes should not be underestimated. Although the usual emphasis in the literature on ToM is on when children have mental state concepts, much of the benefit of having these concepts will come from the ability to use them accurately, and sufficiently quickly to guide fast-moving problems of social interaction and communication. By providing evidence of age-related improvements in speed and accuracy the current study therefore suggests an important contributing factor to continuing improvements in ToM in children who have passed benchmark tests.

Importantly, however, we found little evidence that age-related increases in executive capacity affect the relative difficulty of different belief–desire reasoning problems. There was no age-related change in the effect of true versus false belief: although the absolute error rates and RTs decreased with age, they did not decrease any more for false belief trials than for true belief trials. The pattern for desire was less clear. The effect of positive versus negative desire was present in every age group, but it also interacted with age. However, while the analysis of error rates showed clear evidence of a reduced difference between D+ and D− trials with age, analysis of RTs suggested that this may have come, at least partially, at the cost of the 10- to 11-year-olds trading increased accuracy on D− trials for reduced speed. So it is unclear whether there was really any change in the relative cost of D+ versus D− trials. Altogether, there was little evidence that development consisted in closing the gap in difficulty between harder and easier belief–desire problems.

One possibility is that reasoning with false belief rather than true beliefs, or judging according to negative desires rather than positive desires is a processing threshold, and once a child has the inhibitory capacity to exceed that threshold, additional inhibitory capacity is of no further assistance to the reasoning process. This is significant for conceptualizing the nature of late developmental changes if we take the relative difficulty of B− versus B+ trials as an index of participants' tendency for egocentrism. Consistent with the suggestions of some other authors (e.g., Birch & Bloom, 2007; Epley et al., 2004; Mitchell, Robinson, Isaacs, & Nye, 1996), the current findings suggest a decrease in the magnitude of this tendency for egocentrism over developmental time. But because this decrease was directly proportionate to the overall decrease in speed and error rate, the current findings suggest that it is not egocentrism per se that is decreasing. Instead, the underlying driver of developmental change in ToM performance is a decrease in the absolute processing cost of ToM judgements, which was equally apparent on all of the ToM problems studied here.

Summary

The current findings provided clear evidence of quantitative developmental change in belief–desire reasoning. Among children who would normally pass developmentally sensitive tasks there are significant age-related increases in speed and accuracy of belief–desire reasoning that are likely to result in more efficient ToM performance in everyday settings. However, the current findings also provided clear evidence of qualitative continuity in belief–desire reasoning. The pattern of relative difficulty of different belief–desire tasks that is observed in the ages at which young children first pass these tasks was also observed in the processing costs observed when older children and adults perform the same belief–desire tasks. The most parsimonious explanation for this pattern is that the difficulty of belief–desire reasoning tasks is determined, at all ages, by the varying processing costs of different kinds of belief–desire problem, as predicted by both the executive performance and executive competence accounts of development.

References

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adult performance on a non-inferential theory of mind task. *Cognition*, *106*, 1093–1108.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform account of theory of mind development. *Developmental Psychology*, *45*, 190–201.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaistead, 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.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome. *Journal of Developmental and Learning Disorders*, *5*, 47–78.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*, 382–386.

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature*, *9*, 267–277.

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*, 296–312.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*, 187–192.

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

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, *69*, 672–691.

Cassidy, K. W. (1998). Three- and four-year-old children's ability to use desire- and belief-based reasoning. *Cognition*, *66*, B1–B11.

Chandler, M., Boyes, M., & Ball, L. (1990). Relativism and stations of epistemic doubt. *Journal of Experimental Child Psychology*, *50*, 370–395.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, *17*, 1037–1060.

Doherty, M. J. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Hove, UK: Psychology Press.

Dumontheil, I., Apperly, I., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*, 331–338.

Epley, N., Morewedge, C., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social psychology*, *40*, 760–768.

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

Friedman, O., & Leslie, A. M. (2004). Mechanisms of belief–desire reasoning: Inhibition and bias. *Psychological Science*, *15*, 547–552.

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

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalising. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *358*, 459–473.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in ''theory of mind'': Evidence from compromised belief–desire reasoning in old age. *Cognition*, *101*, 129–152.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, *2*, 861–863.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, *7*, 145–171.

Happe, F. (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 Development Disorders*, *24*, 129–154.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ''theory of mind.'' *Trends in Cognitive Science*, *8*, 528–533.

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

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

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*, 1357–1372.

Mitchell, P. (1996). *Acquiring a conception of mind: A review of psychological research and theory*. Hove, UK: Psychology Press.

Mitchell, P., & Lacohee, H. (1991). Children's early understanding of false belief. *Cognition*, *39*, 107–127.

Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition*, *59*, 1–21.

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cognitive Development*, *10*, 467–482.

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development*, *72*, 688–690.

Paus, T., Evansm, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, *2*, 861–863.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Science*, *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.

Perner, J., Zauner, P., & Sprung, M. (2005). What does ''that'' have to do with point of view? Conflicting desires and ''want'' in German. In J. W. Astington & J. Baird (Eds.), *Why language matters for theory of mind* (pp. 220–244). New York: Oxford University Press.

Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526.

Rakoczy, H., Warneken, F., & Tomasello, M. (2007). ''This way,'' ''No! That way!''—3-year-olds know that people can have mutually incompatible desires. *Cognitive Development*, *22*, 47–68.

Reiss, A. L., Abrams, M. T., Singer, H. S., Ross, J. L., & Denckla, M. B. (1996). Brain development, gender and IQ in children. A volumetric imaging study. *Brain*, *119*, 1763–1774.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*, 12–21.

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

Russell, J. (1999). Cognitive development as an executive process—In part: A homeopathic dose of Piaget. *Developmental Science*, *2*, 247–270.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The ''windows task'' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, *9*, 331–349.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs and false signs. *Child Development*, *77*, 1034–1049.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87–124.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*, 1–12.

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*, 239–277.

Wellman, H. M., Cross, D., & Watson, J. K. (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.

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

Yazdi, A. A., German, T., Defeyter, M. A., & Siegal, M. (2006). Competence and performance in belief–desire reasoning across two cultures: The truth, the whole truth and nothing but the truth about false belief? *Cognition*, *100*, 343–368.