

Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding

Karen Milligan, Janet Wilde Astington, and Lisa Ain Dack

University of Toronto

Numerous studies show that children's language ability is related to false-belief understanding. However, there is considerable variation in the size of the correlation reported. Using data from 104 studies ($N = 8,891$), this meta-analysis determines the strength of the relation in children under age 7 and examines moderators that may account for the variability across studies—including aspect of language ability assessed, type of false-belief task used, and direction of effect. The results indicate a moderate to large effect size overall that remains significant when age is controlled. Receptive vocabulary measures had weaker relations than measures of general language. Stronger effects were found from earlier language to later false belief than the reverse. Significant differences were not found among types of false-belief task.

Children's theory of mind has been a lively area of research in developmental psychology for the past two decades. Work in the area investigates young children's understanding of themselves and other people as mental beings, that is, as people who have beliefs, desires, emotions, and intentions, and whose actions and interactions can be interpreted and explained by taking account of these mental states. The mental state of belief has been a particular focus of interest, with successful performance on experimental false-belief tasks (Wimmer & Perner, 1983) taken to mark the acquisition of a representational theory of mind (Wellman, 2002). The false-belief task assesses a child's ability to reason about the behavioral consequences of holding a mistaken belief. Typically, by 5 years of age children understand that people represent the world in their minds and understand that these representations determine what a person says or does, even in cases where they are misrepresentations of the actual situation in the world.

Meta-analytic findings support the presence of a consistent developmental progression in children's

false-belief understanding that is evident across various countries and various task manipulations (Wellman, Cross, & Watson, 2001). Although children's performance is facilitated by certain task factors, such as creating the false belief with a motivation to deceive someone, age-related changes are still evident. This leads Wellman et al. to argue that developmental changes in false-belief understanding are not an artifact of particular task manipulations but rather a reflection of genuine conceptual change that occurs during the preschool years.

Nevertheless, there is marked variation in the particular age at which individual children achieve success on false-belief tasks. Even on the same version of the test, some children succeed at 3 years of age and others not until 5 years (Jenkins & Astington, 1996). The fact that there is such individual variation in development has led to exploration of possible correlates of false-belief understanding, including cognitive and social variables, in both typically developing and clinical populations (Repacholi & Slaughter, 2003). Language clearly emerges as an important correlate.

This research was supported by a University of Toronto Connaught Research Fellowship and a grant from the Natural Sciences and Engineering Research Council of Canada to JWA. We thank Leslie Atkinson and Ruth Childs for their guidance in completing the analyses for this study, Jodie A. Baird and Eva Filippova for their assistance with coding, and the Theory of Mind research group at the Ontario Institute for Studies in Education, University of Toronto, for assistance in retrieving studies and comments on earlier drafts of this article. We also thank all of the researchers who provided information to supplement that available in published sources.

Correspondence concerning this article should be addressed to Karen Milligan, The Credit Valley Hospital, Child and Family Services, 2200 Eglinton Avenue W., Mississauga, ON L5M 2N1. Electronic mail may be sent to kmilligan@oise.utoronto.ca.

Relation Between Language and False-Belief Understanding

Significant relations between language measures and children's performance on false-belief tasks have been demonstrated in both typically developing children (e.g., Astington & Jenkins, 1999; Hughes & Dunn, 1998; Jenkins & Astington, 1996; Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003) and in clinical samples, such as children with autism

(Happé, 1995; Tager-Flusberg & Joseph, 2005), children with specific language impairment (Miller, 2001), and deaf children (P. de Villiers, 2005; Peterson & Siegal, 1999; Woolfe, Want, & Siegal, 2002). Without doubt, there is a relation between language and false-belief understanding. However, the import of this relation is hotly debated. On the one hand, some researchers argue that it merely reflects the fact that most false-belief tasks are verbal tasks or the fact that the domain-general cognitive operations that underlie false-belief understanding require language for their implementation (e.g., Bloom & German, 2000; Chandler, Fritz, & Hala, 1989; Fodor, 1992; Frye, Zelazo, & Palfai, 1995). On the other hand, many researchers argue that language plays a causal role in the development of false-belief understanding (e.g., J. de Villiers, 2005; Harris, 2005; Nelson, 2005). However, within the latter group there is much dispute over the precise role that language plays (Astington & Baird, 2005a).

There are two main types of studies that report correlations between language and false belief. On the one hand, researchers who are concerned with other cognitive and/or social correlates of false-belief understanding have assessed language ability, often measuring receptive vocabulary or general language ability, in order to control for the effects of language on the relation of interest. On the other hand, researchers have used different kinds of semantic and syntactic language measures in order to investigate more fully and precisely the nature of the relation between language and false belief and the role that language might play in the development of false-belief understanding. This has resulted in an abundance of data on relations between language abilities and false-belief task performance. The strength of reported correlations varies greatly, with some studies suggesting a strong significant relation and others suggesting a weak, nonsignificant one. Integration of these divergent findings has become a necessary and important task. The present study takes up this task using a meta-analytic approach in order to examine and explain the variability across findings.

Language can be operationalized in many different ways, as seen in the wide variety of language measures used in the studies already cited and in many other studies reported in the literature (see Table 1). This is because language is a complex multifaceted system that is used for social communication and for individual mental representation (Astington & Baird, 2005a), which can be assessed in multiple ways. Language is measured from observations of naturalistic conversation, from standardized inventories, and from performance on language-ability tasks.

An important distinction can be made between the conversational and other measures. In conversational discourse, language use is not an individual process but a joint action among participants (Clark, 1996). As Dunn and Brophy (2005) argue, discourse measures of language are emergent properties of a conversation between interlocutors that depend partly on the nature of the relationship between the interlocutors. Such measures cannot be treated solely as characteristics of an individual child. Relatedly, Nelson (2005) argues that when children first start to talk, language is used simply pragmatically, modeled on adult usage. Obviously, coding schemes for children's conversation exclude immediate repetition of adult usage, but even so, early conversational use may not provide a clear measure of an individual child's language ability independent of a particular conversational partner. Some researchers (e.g., Bartsch & Wellman, 1995; Shatz, Wellman, & Silber, 1983) have addressed this problem by not counting pragmatic usage, although this strategy might exclude a large proportion of a young child's conversation. Lastly, a child simply might not have an opportunity to produce a particular linguistic structure during a period of naturalistic observation even though he or she might be capable of producing it in an elicitation task (J. de Villiers & Pyers, 2002). For all of these reasons, the present meta-analysis includes only studies that measured children's language ability using standardized inventories and/or experimental tasks. However, and most important, we do not dismiss the significance of the role that conversation might play in the development of false-belief understanding, an issue that we will take up in the discussion.

The debate concerning the role that language plays in the development of false-belief understanding concerns both spoken and sign language. As mentioned, a significant relation between language measures and children's performance on false-belief tasks is also found in studies with deaf children, whose abilities are assessed using sign language (P. de Villiers, 2005; Peterson & Siegal, 1999; Woolfe et al., 2002). However, the present meta-analysis excludes studies of clinical populations, including deaf children, not because these studies are less interesting or less important, but in order to keep the study samples as homogeneous as possible and limit the number of variables included in the analyses. For the same reason, we include only studies conducted in English.

A thorough search for studies relevant to our purpose, including both published and unpublished, was completed to ensure that the picture presented is not a biased one. Findings from all relevant studies are integrated in order to determine the real strength

Table 1
Articles Included in Meta-Analysis

Study	N	Language ability measure	Type of false-belief task	<i>r/mean r</i>
Ain (Dack) (2004)	32	Test of Early Language Development – III	Change-of-location Unexpected-identity	.43
Angelopoulos and Moore (2003)	118	Test of Early Language Development – II	Unexpected-identity	.45
Artuso (2000)	66	Wechsler Preschool and Primary Scale of Intelligence – Revised (Vocabulary Subtest)	Change-of-location	.53
Astington (2000)	107	Test of Early Language Development – I	Change-of-location Unexpected identity	.61
Astington and Baird (2005b)—Study 1	44	Clinical Evaluation of Language Fundamentals – Preschool Peabody Picture Vocabulary Test	Change-of-location	.36
Astington and Baird (2005b)—Study 2	35	Clinical Evaluation of Language Fundamentals – Preschool Peabody Picture Vocabulary Test	Change-of-location	.25
Astington and Jenkins (1999)	59	Test of Early Language Development – I	Change-of-location Unexpected-identity	.34
Benigno (2004)	33	Peabody Picture Vocabulary Test – Revised	Change-of-location Unexpected-identity	.46
Bialystok and Senman (2004)—Study 2	90	Peabody Picture Vocabulary Test – Revised	Unexpected-identity	.47
Blair, Granger, and Razza (2005)	162	Peabody Picture Vocabulary Test – III	Change-of-location Unexpected-identity	.31
Bowler, Briskman, Gurvidi, and Fornells-Ambrojo (2005)—Study 1	22	British Picture Vocabulary Scale	Change-of-location	.06
Bowler et al. (2005)—Study 2	25	British Picture Vocabulary Scale	Change-of-location	.30
Browning and Holmes-Lonergan (1998)	32	Bankson Language Test – II	Change-of-location Unexpected-identity	.77
Carlson, Mandell, and Williams (2004)	81	Peabody Picture Vocabulary Test – III	Unexpected-identity	.21
Carlson, Moses, and Claxton (2004)	49	Peabody Picture Vocabulary Test – III	Change-of-location Unexpected-identity	.49
Cassidy, Fineberg, Brown, and Perkins (2005)	72	Test of Early Language Development – II	Change-of-location Unexpected-identity	.57
Cassidy, Werner, Rourke, and Zubernis (2003)	67	Test of Early Language Development – II	Change-of-location Deception Unexpected-identity	.66
Chan (2004)	60	Peabody Picture Vocabulary Test Test of Early Language Development – III	Change-of-location Unexpected-identity	.60
Charman, Ruffman, and Clements (2002)	519	British Picture Vocabulary Scale	Change-of-location	.35
Cheung (2006)	25	Peabody Picture Vocabulary Test – III	Belief-desire Change-of-location Unexpected-identity	.46
Cheung et al. (2004)—Study 1	39	Complementation Language Test of Early Language Development – III	Change-of-location Unexpected-identity	.37
Curenton (2003)	72	Early Screening Inventory Language and Cognition Subscale	Change-of-location Unexpected-identity	.45
de Rosnay and Harris (2002)	51	British Picture Vocabulary Scale	Belief-desire	.68
de Rosnay et al. (2004)—Study 2	75	Test for Reception of Grammar	Belief-desire	.50
de Villiers and Pyers (2002)	28	Memory for Complements	Change-of-location Unexpected-identity	.48
Deák and Enright (in press)	42	Flexible Naming	Unexpected-identity	.36
Deák, Ray, and Brenneman (2003)—Study 1	36	Word Knowledge Test	Unexpected-identity	.56
Deák et al. (2003)—Study 2	64	Peabody Picture Vocabulary Test	Unexpected-identity	.52

Table 1. (Contd)

Study	N	Language ability measure	Type of false-belief task	<i>r/mean r</i>
Doherty (2000)—Study 1	48	Homonym Judgement Synonym Judgement British Picture Vocabulary Scale	Change-of-location	.56
Doherty (2000)—Study 2	24	Homonym Selection Task British Picture Vocabulary Scale	Change-of-location	.51
Fahie and Symons (2003)—Study 1	26	Peabody Picture Vocabulary Test	Belief—desire	.35
Farrar and Maag (2002)	20	Test of Early Language Development—III MacArthur Communicative Development Inventory	Change-of-location Unexpected-identity	.62
Flynn (in press)	39	British Picture Vocabulary Scale	Change-of-location Unexpected-identity	.49
Flynn, O'Malley, and Wood (2004)	19	British Picture Vocabulary Scale	Change-of-location Unexpected-identity	.17
Fonagy, Redfern, and Charman (1997)	77	British Picture Vocabulary Scale	Belief—desire	.28
Gal Endres (2003)	54	Stanford Binet Intelligence Scales (Vocabulary Subtest)	Change-of-location Unexpected-identity	.35
Garner, Curenton, and Taylor (2005)	113	Peabody Picture Vocabulary Test—Revised	Change-of-location	.22
Garnham, Brooks, Garnham, and Ostenfeld (2000)—Study 1	43	Homonym Task British Picture Vocabulary Scale	Change-of-location	.49
Garnham et al. (2000)—Study 2	40	Homonyms Task, Synonyms Task, British Picture Vocabulary Scale	Change-of-location	.26
Greig and Howe (2001)	45	British Picture Vocabulary Scale	Unexpected-identity	.26
Guajardo and Turley-Ames (2003)—Study 1	81	Test for the Auditory Comprehension of Language	Deception Unexpected-identity	.61
Guajardo and Turley-Ames (2003)—Study 2	103	Test for the Auditory Comprehension of Language	Deception Unexpected-identity	.64
Hala, Hug, and Henderson (2003)	48	Peabody Picture Vocabulary Test	Change-of-location Unexpected-identity	.24
Hendry (2005)	89	Comprehensive Receptive and Expressive Vocabulary Test—2	Change-of-location Unexpected-identity	.33
Henseler (2000)	50	Test of Early Language Development—II	Unexpected-identity	.38
Holmes (2002)	24	Clinical Evaluation of Language Fundamentals—Preschool—Linguistic Concepts Subtest	Change-of-location Unexpected-identity Other	.38
Hughes (1998)	50	British Picture Vocabulary Scale	Change-of-location Deception Unexpected-identity	.34
Hughes and Cutting (1999)	230	Stanford Binet Intelligence Scales (Vocabulary Subtest and Comprehension Subtests) (composite score used)	Belief-desire Change-of-location Deception Unexpected-identity	.43
Hughes, Dunn, and White (1998)	80	British Picture Vocabulary Scale	Belief-desire Deception Unexpected-identity	.38
Hughes, Jaffee, Happe, Taylor, Caspi, and Moffitt (2005)	2,232	Wechsler Preschool and Primary Scale of Intelligence—Revised (Vocabulary Subtest)	Belief-desire Change-of-location Unexpected-identity Other	.46
Jackson (2001)	48	British Picture Vocabulary Scale	Change-of-location Unexpected-identity Other	.30
James (2002)	53	Peabody Picture Vocabulary Test—Revised	Change-of-location Unexpected-identity	.64
Jenkins and Astington (1996)	68	Test of Early Language Development—I	Change-of-location Unexpected-identity	.64

Table 1. (Contd)

Study	N	Language ability measure	Type of false-belief task	<i>r/mean r</i>
Joe (2003)—Study 1	54	Test of Early Language Development – II	Change-of-location	.56
Joe (2003)—Study 2	75	Preschool Language Scale – 3	Change-of-location	.65
			Unexpected-identity	
Kamawar (2000)—Study 1	73	Stanford Binet Intelligence Scales (Vocabulary Subtest)	Change-of-location	.46
			Unexpected-identity	
Kamawar (2000)—Study 2	64	Stanford Binet Intelligence Scales (Vocabulary Subtest)	Change-of-location	.56
			Unexpected-identity	
Kamawar and de Villiers (2002)	41	Test of Early Language Development – I	Change-of-location	.63
		Complement Comprehension Task	Unexpected-identity	
Keenan (1998)	60	Test of Early Language Development – I	Change-of-location	.30
Liebman (2005)	60	Comprehensive Receptive and Expressive Vocabulary Test – 2	Unexpected-identity	.41
Maring (2003)—Study 1	26	Peabody Picture Vocabulary Test – III	Unexpected-identity	– .19
Maring (2003)—Study 2	80	Peabody Picture Vocabulary Test – III	Change-of-location	.05
			Unexpected-identity	
Mathews, Dissanayake, and Pratt (2003)	78	Peabody Picture Vocabulary Test – Revised	Change-of-location	.17
			Unexpected-identity	
McNab (2001)	80	Test of Early Language Development – II	Change-of-location	.36
Meins, Fernyhough, Russell, and Clark-Carter (1998)	33	British Picture Vocabulary Scale	Change-of-location	.36
			Belief-desire	
Meins, Fernyhough, Wainwright, Das Gupta, Fradley, and Tuckey (2002)	57	British Picture Vocabulary Scale – II	Change-of-location	.37
Miller (2001)	25	Test for the Auditory Comprehension of Language	Change-of-location	–.15
		Reynell Developmental Language Scales – Revised		
Müller, Zelazo, and Imrisek (2005)	69	Peabody Picture Vocabulary Test – Revised	Unexpected-identity	.37
Mullins (2004)	60	Complement Comprehension Task	Change-of-location	.66
			Unexpected-identity	
Murray, Woolgar, Briers, and Hipwell (1999)	94	McCarthy Scales of Children's Abilities	Change-of-location	.24
			Deception	
Nash (2001)	44	Test for the Auditory Comprehension of Language – Revised	Change-of-location	.45
			Unexpected-identity	
Nelson, Plesa, Goldman, Henseler, Presler, and Walkenfeld (2003)	24	Peabody Picture Vocabulary Test	Unexpected-identity	.44
Pears and Fisher (2005)	86	Wechsler Preschool and Primary Scale of Intelligence – Revised	Unexpected-identity	.30
Pears and Moses (2003)	142	Bayley Scales of Infant Development – II or Wechsler Preschool and Primary Scale of Intelligence – Revised	Change-of-location	.13
Pellicano, Maybery, and Durkin (2005)	70	Peabody Picture Vocabulary Test – III	Change-of-location	.44
			Unexpected-identity	
Peterson (2000)—Study 1	98	Peabody Picture Vocabulary Test	Change-of-location	.01
Peterson (2000)—Study 2	167	Peabody Picture Vocabulary Test	Change-of-location	.31
			Unexpected-identity	
Peterson (2004)	15	Peabody Picture Vocabulary Test – Revised	Change-of-location	.03
			Unexpected-identity	
Peterson and Slaughter (2003)—Study 1	61	Peabody Picture Vocabulary Test	Change-of-location	.19
Peterson and Slaughter (2003)—Study 2	47	Peabody Picture Vocabulary Test	Belief-desire	.45
Razza (2005)	73	Peabody Picture Vocabulary Test – III	Change-of-location	.31
			Unexpected-identity	
Renouf et al. (2006)	66	PPVT	Unexpected-identity	.17
Repacholi and Trapolini (2004)	48	Peabody Picture Vocabulary Test – Revised	Unexpected-identity	.42

Table 1. (Contd)

Study	N	Language ability measure	Type of false-belief task	<i>r/mean r</i>
Ruffman, Perner, and Parkin (1999)	64	British Picture Vocabulary Scale	Change-of-location	.00
Ruffman, Slade, Clements, and Import (1998)—Study 2	64	Clinical Evaluation of Language Fundamentals—Preschool (Linguistic Concepts and Sentence Structure Subtests)	Change-of-location Unexpected-identity	.19
Ruffman, Slade, and Crowe (2002)	79	British Picture Vocabulary Scale Clinical Evaluation of Language Fundamentals—Preschool (Linguistic Concepts Subtest)	Change-of-location Other	.60
Ruffman et al. (2001)—Study 1	59	Clinical Evaluation of Language Fundamentals—Revised (Sentence Structure Subtest)	Change-of-location Unexpected-identity Other	.14
Ruffman et al. (2001)—Study 2	65	British Picture Vocabulary Scale Embedded Syntax Task Word Order Syntax Task	Change-of-location Unexpected-identity	.39
Schwebel, Rosen, and Singer (1999)	54	Peabody Picture Vocabulary Test—Revised	Change-of-location Unexpected-identity	.47
Seamans (2004)	37	Peabody Picture Vocabulary Test—III	Unexpected-identity	.30
Slade and Ruffman (2005)	44	British Picture Vocabulary Scale Clinical Evaluation of Language Fundamentals (Linguistic Concepts Subtest) Embedded Clause Test Word Order Test	Change-of-location Unexpected-identity	.59
Smith, Apperly, and White (2003)	56	Relative Clause Sentences	Change-of-location Unexpected-identity	.54
Symons, Peterson, Slaughter, Roche, and Doyle (2003)—Study 1	51	Peabody Picture Vocabulary Test—III	Belief-desire Change-of-location Unexpected-identity	.29
Symons et al. (2003)—Study 3	20	Peabody Picture Vocabulary Test—III	Belief-desire Change-of-location Unexpected-identity	.16
Taylor and Carlson (1997)	150	Peabody Picture Vocabulary Test	Unexpected-identity	.28
Watson (2000)	71	Peabody Picture Vocabulary Test—Revised	Change-of-location Unexpected-identity	.27
Watson, Nixon, Wilson, and Capage (1999)—Study 1	26	Test for the Auditory Comprehension of Language—Revised	Unexpected-identity	.32
Watson et al. (1999)—Study 2	52	Test for the Auditory Comprehension of Language—Revised	Change-of-location Unexpected-identity	.72
Watson, Painter, and Bornstein (2001)	36	MacArthur Communicative Development Inventory—Toddlers Reynell Developmental Language Scales—Revised	Unexpected-identity	.56
Weimer and Guajardo (2005)	60	Test for the Auditory Comprehension of Language—Revised	Change-of-location; Deception Unexpected-Identity	.42
Wellman, Phillips, Dunphy-Lelii, and LaLonde (2004)	17	Peabody Picture Vocabulary Test	Change-of-Location; Unexpected-Identity	.23
Whitehouse and Hird (2004)	23	Clinical Evaluation of Language Fundamentals—Preschool or Clinical Evaluation of Language Fundamentals—III	Change-of-Location	.88
Wilson (1998)	63	Test for the Auditory Comprehension of Language—Revised	Unexpected-Identity	.27
Ziatas, Durkin, and Pratt (1998)	24	Peabody Picture Vocabulary Test; Test of Reception of Grammar	Change-of-Location	.40

of the relation between language ability and false-belief understanding. Then we examine potential moderators of this relation.

Potential Moderating Factors

In addition to the potential effects of general demographic and study quality variables, we examine the potential effects of the type of language ability assessed, the type of false-belief test used, and the direction of effect in longitudinal studies on the strength of the relation between language ability and false-belief understanding. These latter three factors are of substantial interest in any discussion of the role that language plays in the development of false-belief understanding.

Type of language ability. As already mentioned, language is a complex, multifaceted system. Consequently, an important question is whether all or only particular aspects of the linguistic system are involved in the relation with false-belief understanding. A further question is whether some aspects of language are more strongly related to false-belief understanding than others. Studies drawn on for the meta-analysis used a wide variety of language measures, allowing us to address these questions, and thus to contribute to the debate on the role that language might play in the development of false-belief understanding.

Language abilities include semantics and syntax, as well as pragmatics. Pragmatic ability underlies the capacity to use and interpret language appropriately in communicative exchanges. This ability is generally assessed by measures extracted from naturalistic conversation and therefore it is not considered here because, as mentioned, such measures are inherently dyadic in nature. In the present meta-analysis we focus on the individual child's abilities assessed by standardized inventories and experimental measures. Naturally these abilities will be partly dependent on the child's social linguistic environment but they are individual cognitive measures in a way that is clearly different from dyadic measures.

Semantic ability consists of both lexical (i.e., word) knowledge and aspects of meaning that go beyond the word level, that is, discourse semantics. Lexical knowledge can be assessed from measures of receptive vocabulary, such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997) and the related British Picture Vocabulary Scale (Dunn, Dunn, Whetton, & Burley, 1997). Other semantic tests go beyond the single-word level, such as the Linguistic Concepts subtest in the Clinical Evaluation of Lan-

guage Fundamentals–Preschool (CELF–P) (Wiig, Secord, & Semel, 1992).

Syntactic ability involves mastery of linguistic structure, whereby words are combined into sentences. It is measured by tests such as the Sentence Structure subtest of the CELF–P (Wiig et al., 1992). Other tests focus on the mastery of particular syntactic structures, such as sentential complements, assessed by the memory for complements test (J. de Villiers & Pyers, 2002). A sentential complement is a tensed subordinate clause that is embedded under a mental or communication verb to form a complex sentence, for example: Maxi thinks [*that*] the chocolate is in the cupboard (complement italicized; the specific complementizer “that” is optional). This construction allows for a true report of a mistaken representation—because the whole complex sentence, consisting of main and embedded clauses, can be true even though the embedded clause expresses a proposition that is false. That is, this syntactic structure provides the format needed to represent false beliefs.

General language measures (e.g., the Test of Early Language Development, TELD–III, Hresko, Reid, & Hammill, 1999) include a combination of semantic and syntactic items. Semantic and syntactic abilities covary in typically developing children, even though they can be assessed separately. Moreover, some tests simultaneously draw on more than one ability even though a specific ability is the focus of the test; for example, children require semantic ability to comprehend the words used in syntax tests.

When considering potential moderators in the present meta-analysis, five types of language ability are examined: general language, semantics, receptive vocabulary, syntax, and memory for complements. This allows us to address some of the disagreements that currently exist in the literature concerning the particular importance of specific aspects of language ability in the development of false-belief understanding. For instance, studies examining the roles of syntax and semantics have produced inconsistent results. Astington and Jenkins (1999) argued that syntactic ability was particularly important to false-belief understanding, based on their finding that syntax but not semantics predicted unique variance in false-belief task scores. In contrast, Ruffman et al. (2003), using a different measure of syntax and semantics, reported the converse finding, that is, semantics but not syntax predicted unique variance in false-belief task scores. In a similar way, studies examining the roles of general language ability and the ability to understand sentential complements have also produced conflicting results. In one study (J. de Villiers & Pyers, 2002), memory for comple-

ments made a unique contribution to false-belief understanding beyond that of more general language measures, whereas in another study (Cheung et al., 2004), a somewhat different measure of memory for complements did not make a unique contribution to false-belief understanding beyond that of general language ability.

Recently, there have been numerous studies like these just cited that examine the role of different aspects of language in the development of false belief understanding. This research aims to provide evidence for theoretical explanations of the development of false-belief understanding, in particular, and of theory of mind more generally conceived (Astington & Baird, 2005a). On the one hand, language is thought to be important because it allows the child to participate fully in the culture, to engage in social interaction, to join in conversation, to listen to stories, and so on (Nelson, 2005). And, it is argued, that it is these activities that promote the development of theory of mind, including false-belief understanding. On this perspective, one would predict that general language measures, or measures of semantics or vocabulary would relate to false-belief task performance because these measures provide a good indication of the child's ability to participate in linguistic social interaction. On the other hand, language is thought to be important because it provides the child with representational resources for managing false beliefs. On this view, one would predict that measures of syntax (Astington & Jenkins, 1999) and memory for complements (J. de Villiers, 2005) would relate to false-belief task performance. Syntactic ability enables the child to keep track of changing locations in false-belief stories, and the ability to use sentential complement constructions allows the child to represent false beliefs.

Findings from the meta-analysis cannot be expected to resolve the debate about causal mechanisms but they can make an important contribution. For instance, the results may provide more support for one explanation over another. They can also provide detailed guidance toward the kind of studies that are needed to answer the outstanding questions.

Type of false-belief task. Wellman et al. (2001) argued that conceptual change, independent of task factors, underlies children's false-belief task performance. This does not, however, rule out the possibility that performance on different types of task may differ by language ability because of the different linguistic demands of false-belief tasks. The present study therefore examines the strength of the relation between false-belief understanding and language ability for different types of false-belief measure.

Studies drawn on for the meta-analysis used a range of false-belief tasks, including the widely known and frequently used standard change-of-location and unexpected-identity tasks, as well as deception tasks and belief-based emotion-attribution tasks.

In the change-of-location false-belief task (Wimmer & Perner, 1983) children have to follow the details of a narrative in which an object is moved from one location to another while the story protagonist is off the scene (e.g., Maxi's chocolate is moved from a cupboard to a drawer). When the protagonist returns to the scene, children are asked where he thinks the object is, or more simply, where he will look for it (e.g., "Where does Maxi think the chocolate is?" or "Where will Maxi look for the chocolate?"). The unexpected-identity false-belief task (Perner, Leekam, & Wimmer, 1987) was designed to reduce task demands by dispensing with the narrative and giving children themselves the experience of holding a false belief. In this task, children are shown a familiar container, asked what is inside, and then it is opened to reveal atypical, unexpected contents (e.g., pencils inside a Smarties candy box). The child is asked what another person will think is inside it, when first seeing the closed box (e.g., "What will your friend think is in the box?"). Sometimes the child is also asked what he or she thought was inside the box before it was opened. In a variant of the unexpected-identity task (Gopnik & Astington, 1988), developed from Flavell's (1986) appearance-reality task, children are shown an object or picture that looks like one thing but is really something else, and the child is asked what another person will think the object or picture is, when first seeing it before the true identity is revealed (e.g., an apparent "rock" is really a sponge, or what appear to be a cat's ears when only part of a picture is in view are seen to be flower petals when the whole picture is revealed).

Some type of change-of-location and/or unexpected-identity false-belief task is used in almost all studies of the relation between language and false-belief understanding, whereas deception and belief-based emotion-attribution tasks are used less frequently. Deception tasks require that the child actively deceive the experimenter or a puppet character, for example, by concealing an object or by saying something that is false. The deception tasks used in studies included in the meta-analysis vary considerably in their linguistic demands. In some tasks the child has to follow a narrative as in the change-in-location task and then predict what the protagonist would *say* if he or she wanted to deceive another character about the location of the object (Guajardo & Turley-Ames, 2004). Other deception tasks are much less verbally

demanding, such as a penny-hiding game, which the experimenter models before asking the child him- or herself to hide the penny (Hughes, 1998). Belief-based emotion-attribution tasks (Harris, Johnson, Hutton, Andrews, & Cooke, 1989) also require the child to follow a narrative and then ask the child to predict a character's emotion in a situation where the character holds a false belief. In order to make the correct prediction in these tasks, the child has to assess whether characters believe their desires will be fulfilled, not simply whether the desires will be fulfilled. However, the test question is quite simple ("How does [the character] feel?")

In summary, false-belief tasks used in studies included in the meta-analysis impose a variety of linguistic demands on children. First, change-of-location, belief-emotion, and some deception tasks involve a narrative whereas other tasks do not (i.e., unexpected-change and some deception tasks). Second, unexpected-change, some change-of-location, and some deception tasks involve a test question with an embedded sentential complement whereas other tasks use a simpler test question (i.e., belief-emotion, some change-of-location, and some deception tasks). Hence, it is possible that performance on the different types of task may differ by language ability because of the different linguistic demands.

Direction of effect. The relation between language ability and false-belief understanding might be explained in a number of different ways, depending on whether language predicts false belief over time, or false belief predicts language, or each predicts the other in a bidirectional relation. First, as mentioned above, many researchers make a case for a predictive relation from language to false belief, arguing that language plays a causal role in the development of false-belief understanding (Astington & Baird, 2005a). Data from longitudinal studies (Astington & Jenkins, 1999; J. de Villiers & Pyers, 2002) and training studies (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003) support these arguments. Second, however, it is possible that the causal relation might operate in the other direction, that is, the development of false-belief understanding promotes language development. This argument is supported by the fact that language is not necessary for false-belief task performance because adults who lose the capacity for language are still able to perform correctly on false-belief tasks (Varley, 1998; Varley, Siegal, & Want, 2001). Yet third, there may be a bidirectional relation between false-belief understanding and language ability because each promotes development of the other, either by way of a bootstrapping mechanism (Shatz, 1994) or because

they are both related to some other underlying factor, such as executive functions (Carlson & Moses, 2001).

Studies drawn on for the meta-analysis measured language ability and false-belief understanding either concurrently or longitudinally; in the latter case the false-belief assessment either preceded or followed the language assessment. We examine the strength of the relation between false belief and language in studies where the false-belief assessment preceded the language assessment by 3 months or more, and/or the false-belief assessment followed the language assessment by 3 months or more. This allows us to compare the predictive strength of the relation from earlier language ability to later false-belief understanding, and from earlier false-belief understanding to later language ability, in order to examine support for arguments concerning the direction of effect of language ability on false-belief understanding.

Summary of Aims

In summary, the present meta-analysis has two main aims. The first aim is to integrate findings across a large number of studies in order to determine the real strength of the relation between language ability and false-belief understanding. The second aim is to examine the potential moderators of this relation by performing a number of more detailed secondary analyses. In addition, because language ability and false-belief understanding are both related to age, a further secondary analysis is conducted in order to determine the strength of the relation when the influence of age is taken into account. Taken all together, these analyses allow us to contribute to the debate on the role of language in the development of false-belief understanding and, more generally, to address the question of why language matters for theory of mind.

Method

Retrieval of Studies

An extensive search for studies addressing the relation between false-belief task performance and language ability was conducted. First, databases, including *Dissertation Abstracts International*, *Linguistics and Language Behavior Abstracts*, *MedLine*, and *PsycINFO*, were searched for relevant studies published after 1980. The keywords "theory of mind," "false belief," "mental representations," "mental states," and "belief-desire reasoning" were used as major descriptors on their own and in combination with "language ability." Abstracts of studies were

then read to determine broad relevance (e.g., empirical study, children served as participants, language of study, false-belief task administered). Given that language ability often served as a covariate in studies and was not the primary focus of research, all studies using false-belief tasks were searched to determine whether a language measure was also administered. Second, relevant journals in the area of child development (e.g., *British Journal of Developmental Psychology*, *Child Development*, *Cognitive Development*, *Current Psychology of Cognition*, *Developmental Psychology*, *International Journal of Language and Communication Disorders*, *Journal of Child Language*, *Journal of Child Psychology and Psychiatry*, *Journal of Cognition and Development*, *Journal of Communication Disorders*, *Mind & Language*, and *Social Development*) were searched manually. Third, conference proceedings from the Biennial Meeting of the Society for Research in Child Development (2003 and 2005) and the Annual Meeting of the Jean Piaget Society (2004 and 2005) were searched using the keywords listed above. Given that abstracts were generally not available, conference proceedings were used as a means of identifying researchers in the area. These researchers were contacted and relevant published or unpublished data were requested. Finally, we contacted all individuals who attended the International Conference on Why Language Matters for Theory of Mind (held at the University of Toronto in April 2002), as well as additional researchers in the area of theory of mind known to the authors, and requested any published or unpublished data that addressed the relation between false-belief task performance and language ability. In order to ensure that the data were representative and not skewed in favor of positive results, particular emphasis was placed on searching for "fugitive data," which includes data that are difficult to obtain through traditional sources, including unpublished data, unsubmitted/rejected papers and manuscripts, and dissertations and theses (Rosenthal, 1994). Our efforts to find relevant studies were concluded in February 2006. Data were included from some unpublished manuscripts that have since been published (e.g., Slade & Ruffman, 2003, published as Slade & Ruffman, 2005). These studies and their published reference are noted in the reference list. Given the size of the literature, it is possible that the studies included in the present meta-analysis do not represent an exhaustive list. However, given the breadth of the search methods and sources used, the studies included are considered to be well representative of the published and unpublished data in the area.

Table 2
Type of False-belief Task

Type of task	Examples of tasks included
Change-of-location	Maxi task Sally-Ann task
Unexpected-identity	Appearance reality Misleading appearance Smarties test Unexpected change Unexpected contents Unexpected picture
Deception task	Modified unexpected change task One-box deception game Penny hiding game
Belief-emotion	Belief-based emotion attribution Emotion false-belief task Nice/nasty surprise
Other	Any other task, other than a first-order false-belief task, included in false-belief composite score (e.g., emotion understanding task; false-belief explanation task; second-order false-belief task)

Inclusion and Exclusion Criteria

The search procedures described generated 324 studies. Stringent inclusion/exclusion criteria were used. (1) This meta-analysis was limited to studies of the relation between false belief and language in typically developing, English-speaking children. If data examining the relation between false-belief task performance and language ability were available for a typically developing control group in the context of a study addressing the relation for a clinical population, data for the control group were included. Samples that included children with behavior problems and learning disabilities were included as these characteristics may be present and not screened for in a study including typically developing children. (2) Children had to be <7 years of age. (3) A standardized and/or experimental measure of language ability (see Table 1) had to be administered to children. If language was assessed solely through the use of measures of naturalistic conversation (e.g., talkativeness, mental state terms) or verbal reasoning (e.g., comprehension or information subtests of the Wechsler Intelligence Scale for Children), the study was not included. (4) Studies had to include at least one first-order false-belief task (see Table 2). Studies that administered only emotion-understanding tasks, false-belief explanation tasks, second-order false-belief tasks, or other perspective-taking tasks that did not

include a false-belief element were not included. (5) Training studies or studies examining children who did not pass false-belief tasks were not included because there was not adequate range in false-belief task performance. (6) Only studies conducted in English were included. The absence of a measure of language ability accounted for the most exclusions. Altogether, 104 studies with a combined sample size of 8,891 met the inclusion criteria (see Table 1).

Studies were coded for inclusion/exclusion by K. M. and by one of two independent PhD level research assistants, J. A. B. and E. F., who were not involved in retrieving the studies or any other aspect of this meta-analysis. Interrater reliability was calculated using κ statistics and indicated a high level of agreement for inclusion of studies between coders (.91).

Coding of Study Features

A detailed coding system was developed to evaluate each study meeting inclusion criteria. Studies were coded on the following features:

1. Year of publication;
2. Source of study (i.e., journal, dissertation, conference paper, book, unpublished data);
3. Sample size used for correlation between language ability and false-belief task performance (if sample size differed across correlations that were to be combined for an overall correlation, the average sample size was used);
4. Mean age of participants in months;
5. Male/female ratio in the sample;
6. Language task used (see Table 1);
7. Type of false-belief task used (see Table 2);
8. Number of different types of false-belief task;
9. Number of trials of false-belief task;
10. Inclusion of control questions;
11. Direction of effect (i.e., false-belief assessment preceded the language assessment by 3 months or more, or false-belief assessment followed the language assessment by 3 months or more);
12. Correlation coefficient for the relation between false-belief task performance and language ability; and
13. Partial correlation coefficient for the relation between false-belief and language tasks, controlling for age (where available).

Studies that met inclusion criteria were coded for the above study features by K. M. Twenty percent of studies were randomly selected and re-coded by L. A. D to establish inter-rater reliability. For categorical variables (e.g., type of false-belief task, inclusion of control questions), Cohen's $\kappa = 1.0$,

indicating perfect agreement between coders. For continuous variables (e.g., mean age, male/female ratio, number of false-belief task trials), agreement between coders ranged from 90% to 100%. All disagreements were discussed and a consensus was reached between the two coders. If studies did not report sufficient data for coding the study characteristics and results, researchers were contacted and the missing data were requested. Fifty-eight researchers were contacted. Thirty-seven (64%) provided the requested data and these data were included in the analyses. Data from 12 researchers who responded to the request (21%) were not included because their study did not meet inclusion criteria or the data requested were unavailable. Seven researchers (12%) did not respond to the request for data and 2 researchers (3%) could not be located.

Data Analyses

Pearson product-moment correlation coefficients (r) from each study were used to assess the strength of the relation between language ability and false-belief task performance across studies. In situations in which multiple studies presented findings based on the same sample (or subset) of children or presented correlations between multiple measures of language ability and/or false-belief understanding, a single effect size was calculated, as suggested by Rosenthal (1991). This was done in order to ensure that results from studies were not disproportionately weighted and the strength of the observed effect was not biased. More specifically, the following procedures were used. (1) If data collected from the same sample of children using the same measures of language ability and false-belief understanding were published in multiple articles, results from the most recent study and/or the study presenting the correlations between language ability and false-belief task performance were used. (2) If studies were based on the same measures of language ability and false-belief understanding but reported on different subsamples of the same group of children, results from the study with the largest sample were used. (3) If studies with the same sample of children presented data on different measures of language ability, the correlations between each measure of language ability and false-belief understanding were averaged across studies to compute a single mean effect size (i.e., the sum of the Fisher Z_r transformed scores weighted by sample size, divided by overall sample size). (4) If correlations between multiple measures of language ability and false-belief understanding were reported, correlation coefficients were

averaged across measures to compute a single mean effect size.

Comprehensive Meta-Analysis, Version II (Borenstein, Hedges, Higgins, & Rothstein, 2005) was used to calculate the overall effect size between language ability and false-belief understanding and secondary relations of interest (i.e., type of language ability, type of false-belief task, direction of effect) using standardized scores (Fisher Z_r scores) weighted by inverse variance. Only studies that provided separate data for these relations were included. Random effects models were used to calculate effect sizes if Cochran's chi-square tests for heterogeneity were significant. If significant heterogeneity was not indicated, fixed effects models were used. Random effects models have been argued to represent the heterogeneity of behavioral studies more adequately than fixed effect models. Noninflated alpha levels are used when the requirement of homogeneity has not been met (Hunter & Schmidt, 2000). Ninety-five percent confidence intervals (CI) for the effect size are also presented.

Results

Overall Relation Between Language Ability and False-belief Understanding

The overall effect size weighted by inverse variance for the relation between language ability and false-belief task performance using a random effects model was .43 (CI = .39–.47; SE = .02), which, according to Cohen (1988), is moderate to large in strength, and is significantly greater than zero ($z = 21.12$, $p < .001$). The file drawer statistic was also calculated to assess the reliability of this finding. This represents the number of unretrieved studies averaging null results that would be required to reduce the significance of the meta-analytic finding to just the significant level, $p = .05$ (Rosenthal, 1991). Results indicated that 31,982 studies with null results would be needed to reduce results to the just significant level, which well exceeds Rosenthal's critical value of 530 ($5k + 10$, where k is the number of studies). We can therefore be confident that this significant result would not be negated by null findings that were not included in the present analysis.

Given that the sample size of one of the included studies was significantly different from the other studies ($n = 2,232$; Hughes et al., 2005), the overall effect size weighted by sample size was also calculated with this study omitted. Based on 103 studies ($n = 6,659$), the overall weighted effect size using a

Table 3
Effect Sizes for Relation Between Language Ability and False-belief Task Performance

Stem	Leaf
–.1	95
.0	01356
.1	3467799
.2	1234456677889
.3	000000111234455566667778889
.4	0122334455556666778999
.5	01245666679
.6	0011234445668
.7	27
.8	8

random effects model was .43 (CI = .39–.47; SE = .02), which is significantly greater than zero ($z = 19.75$, $p < .001$). Given that the effect size did not significantly differ, all 104 studies were included in the examination of general demographic characteristics and study quality variables that may play a moderating role. This study did not report data relevant to the secondary analyses (i.e., correlations for specific type of language ability, specific false-belief tasks, and direction of effect) and was therefore not involved in the calculation of other effect sizes discussed.

A stem and leaf plot illustrating the distribution of effect sizes is presented in Table 3. This table shows the actual effect size for each one of the 104 studies included in the analysis, ranging from $r = -.19$ to .88. As illustrated in Table 3, language accounts for between 7% and 25% of the variance in false-belief understanding for the two central quartiles of the studies. However, in the upper quartile, the proportion of variance accounted for ranges from 26% upward to 77%, and in the lower quartile it ranges from 7% down to below zero. Cochran's chi-square test was completed to examine the homogeneity of variance and indicated that effect sizes significantly differed across studies, $Q = 290.81$, $p < .001$. Given that skew (-0.33 ; SE = .24) and kurtosis (.59; SE = .47) statistics were within acceptable limits, parametric statistics were used to look at potential moderators that might explain the differences found among studies.

Moderators of the Relation Between Language Ability and False-belief Understanding

Two types of analyses were completed to examine the potential moderators of the relation between

language ability and false-belief task performance. First, general demographic and study quality variables were examined using the data from the overall results described above. Second, separate secondary meta-analyses were conducted to look more specifically at the moderating role of type of language ability assessed, type of false-belief task used, and direction of effect.

General demographic characteristics and study quality variables. Two types of moderators were explored: general demographic characteristics and study quality. Analyses revealed that study effect size did not significantly differ by male/female ratio in the sample, mean age of participants, source of study, or year of publication. With regard to study quality, studies including a greater number of types of false-belief tasks reported higher correlations between language ability and false-belief performance ($r = .19$, $p = .05$). However, studies did not significantly differ by number of false-belief task trials ($r = .11$, ns) or by inclusion of control questions $t(1, 101) = .80$, ns . Effect sizes also did not significantly differ by sample size ($r = .03$, ns).

Type of language ability. Given that some measures assessed general aspects of language ability and other measures assessed specific aspects, such as semantics (e.g., sentence meaning, synonyms, homonyms), receptive vocabulary, syntax (e.g., word order, sentence structure), and memory for complements, analyses were completed to determine whether the strength of the relation between language ability and false-belief task performance differed by type of language ability assessed. Five types of language ability (i.e., general language, semantics, receptive vocabulary, syntax, and memory for complements) were examined. Only the most common measures of general language and receptive vocabulary in the included studies were examined in secondary analyses. All standardized and/or experimental measures of semantics, syntax, and memory for complements were included. See Table 4 for measures included for each type of language ability. Average effect sizes weighted by inverse variance are presented in Table 5. The strength of effect size for type of language measure ranged from .35 (moderate, 12% of the variance) to .66 (strong, 44% of the variance). An ANOVA was completed using mixed effects analyses to determine whether effect size significantly differed by type of language ability. A significant main effect was found for type of language ability, $Q(4, 103) = 25.32$, $p = < .001$. Pair-wise comparisons using mixed effects analyses were completed to examine which types of language ability differed significantly. Given the large number of

Table 4
Type of Language Measure

Language ability assessed	Measure
General language	Test for Auditory Comprehension of Language Test of Early Language Development (I, II, and III)
Semantics	Bankson Language Test – II (Semantic items) Clinical Evaluation of Language Fundamentals – Preschool (Linguistic Concepts subtest) Homonym Judgement task Synonym Judgement task Test of Early Language Development I (Semantic items) Word Order Semantics task
Receptive vocabulary	British Picture Vocabulary Scale Peabody Picture Vocabulary Test
Syntax	Bankson Language Test – II (Syntax items) Clinical Evaluation of Language Fundamentals-Preschool (Sentence Structure subtest) Embedded Clause test MacArthur Communicative Development Inventory (Complexity of grammar items) Test of Early Language Development I (Syntax items) Test of Reception of Grammar Word Order Syntax task
Memory for complements	Complement Comprehension Task Memory for Complements Task

Note. Not all language measures used in studies included in the meta-analysis were coded according to the above scheme. Only those specific measures listed were included in secondary analyses, as these were the most common measures employed.

pair-wise comparisons completed (10), a more conservative significance level of $p = .005$ (Bonferroni correction .05/10) was used. The results of these comparisons indicated that the relation between false-belief performance and performance on receptive vocabulary measures was significantly lower than performance on general language measures. No other significant differences were found.

Type of false-belief task. In order to determine whether the strength of the relation between language ability and false-belief task performance differed by the type of false-belief task administered, the strength of the relation for specific types of false-belief task was examined: (1) change-of-location, (2) unexpected-identity, (3) deception, and (4) belief-emotion. Average effect sizes weighted by inverse variance are presented in Table 6. The strength of

Table 5

Average Effect Size for Relation Between False-belief Task Performance and Different Types of Language Ability Measures

Type of language ability measure	M	k	SE	CI
General language	.52*	23	.04	.44–.60
Receptive vocabulary	.34*	53	.03	.28–.40
Semantics	.48*	13	.06	.36–.60
Syntax	.54*	11	.08	.38–.70
Memory for complements	.66*	4	.08	.51–.82

Note. Random effects model used for all analyses except memory for complements. CI = 95% confidence interval; *k* = number of studies; *M* = mean ES.

* $p < .001$.

effect size for type of false-belief task type ranged from .35 (moderate, 12% of the variance) to .52 (strong, 26% of the variance). An ANOVA using mixed effects analyses indicated that effect size differed significantly by type of false-belief task, $Q(3, 90) = 9.035, p = .03$. Pair-wise comparisons using mixed effects analyses were completed to examine which types of false-belief task differed significantly. Given the number of pair-wise comparisons completed, a more conservative significance level of $p = .008$ (Bonferroni correction $.05/6$) was used. Although the omnibus random effects test across all four types of false-belief task yielded a significant main effect of false-belief task, the results of pair-wise comparisons indicated no significant differences between types of false-belief task.

Direction of effect. To address the question of whether language ability predicts false-belief task performance or false-belief task performance predicts language over developmental time, the strength of the relation for each was examined. The largest average effect size (using the fixed effects model) was found for earlier language ability and later false-belief task performance ($ES = .56, k = 11, CI = .46-.66$;

$SE = .05$), which is significantly greater than zero ($z = 11.28, p < .001$). The effect size for early false-belief task performance and later language test performance ($ES = .36, k = 7, CI = .24-.48; SE = .06$) was also significant, ($z = 5.76, p < .001$). The results of an ANOVA using mixed effects analyses indicated that the relation between earlier language ability and later false-belief task performance was significantly stronger than the relation between earlier false-belief task performance and later language ability, $Q(1, 17) = 5.57, p = .02$.

Language Ability, False-belief Understanding, and Age

Finally, although the results show that the relation between false-belief task performance and language ability is significant and moderate to large in strength, it is possible that the strength of the effect found is due primarily to age, which is known to be strongly related both to false-belief understanding and language ability. To determine the relation between false-belief task performance and language ability, controlling for the influence of age, a secondary meta-analysis was completed using all studies that provided correlations for false-belief task performance and language ability controlling for age. Based on 32 studies ($n = 1,616$), the overall effect size weighted by inverse variance and using a fixed effects model was .31 ($CI = .25-.37; SE = .03$), which is significant ($z = 11.93, p < .001$) and moderate in strength. While a wide range of study effect sizes was indicated ($ES = -.14$ to $.63$), Cochran's chi-square test for heterogeneity indicated that inter-study effect size differences were not significant, $Q = 32.77, ns$. The file drawer statistic indicated that 1,079 studies with null results would be needed to reduce the effect size to just significant, which exceeds Rosenthal's (1991) critical value (170). We can therefore be confident that this significant result would not be negated by null findings that were not included in the present analysis.

Discussion

The meta-analysis presented here clearly demonstrates that there is a significant relation between children's language ability and their false-belief understanding that is independent of age. This finding comes as no surprise, given the number of reports of such a relation that have appeared in the literature in recent years. The meta-analysis, however, adds an important new dimension because it combines results from 104 studies with a total sample size of almost 9,000 children. The overall relation, not con-

Table 6

Average Effect Size for Relation Between Language Ability and Performance on Different False-belief Task Types

Task	M	k	SE	CI
Change-of-location	.35*	40	.03	.51–.82
Unexpected-identity	.42*	37	.001	.42
Deception task	.52*	6	.08	.37–.67
Belief-emotion	.48*	8	.05	.39–.57

Note. Random effects models were used for the Change-of-location and Deception task. Fixed effects models were used for the unexpected-identity and belief-emotion tasks. CI = 95% confidence interval; *k* = number of studies; *M* = mean ES.

* $p < .001$.

trolling for age, is moderate to large in strength ($r = .43$), while in the subset of studies that controlled for the effect of age the effect size remains moderate ($r = .31$). That is, language ability overall accounts for 18% of the variance in false-belief task performance. Moreover, even though language ability strongly covaries with age, it still predicts a highly significant 10% of the variance in false-belief understanding after accounting for the variance attributable to age.

An additional finding is the remarkable variability in the strength of the relation across all of the studies included in the meta-analysis, which may well be due to the variety of measures and tasks used. As illustrated in Table 3 and noted in the results, the proportion of variance in false-belief task performance that language accounts for ranges from below zero upward to an impressive 77%. Even in studies that controlled for the effects of age, the proportion of variance in false-belief task performance that language accounts for ranges from below 0% up to 40%. The findings obviously lead to the important question of what factors affect the variability in the strength of the relation between language ability and false-belief understanding. The analysis shows that this variability is not due to general demographic characteristics of the participants, such as their mean age or the male/female ratio in the sample. Nor is the variability due to characteristics of the studies, such as when the data were reported, and whether they were published or unpublished. Likewise, the variability is not due to several aspects of the quality of the study, including sample size, number of false-belief task trials given, and whether or not control questions were included. Some variability is explained, however, by one aspect of study quality. Specifically, the effect size is significantly greater for studies that included a greater number of types of false-belief task. Certainly, one might expect to find stronger correlations in studies that use multiple measures of one or both of the constructs in the relation because this is likely to increase variability in the scores. It may be that using different types of false-belief tasks maximizes the opportunity for children to reveal their understanding of false belief.

Beyond effects due to the characteristics of the participants and the characteristics of the studies, more interest is undoubtedly directed toward the substantive issue of the potential moderating effects of the type of language ability assessed and the type of false-belief task used, as well as the direction of effects in longitudinal studies. The results of the secondary meta-analyses that were conducted address these questions.

Type of Language Ability

A pressing question in the literature is whether all or only particular aspects of language are involved in the relation with false-belief understanding. In the present meta-analysis, a great variety of language measures was used overall (see Table 1). Five different types were examined in the secondary meta-analysis: general language, semantics, receptive vocabulary, syntax, and memory for complements. The results clearly show that all of these aspects of the linguistic system are involved in the relation with false-belief understanding. That is, there is a significant relation between language ability and false-belief task performance for each type of language ability (see Table 5). Thus, it is clear that false-belief understanding is related to various aspects of language.

A further pressing question is whether some aspects of language are more strongly related to false-belief understanding than others, and here our findings are more equivocal. True, the effect sizes differ by type of language ability, with receptive vocabulary accounting for 12% of the variance in false-belief understanding, semantics accounting for 23%, general language for 27%, syntax for 29%, and memory for complements for 44%. However, partly due to the number of studies included in each category, the only difference that proves to be significant on post hoc pair-wise comparisons is that between receptive vocabulary and general language. This may well be because receptive vocabulary tests can assess an isolated language ability whereas other tests simultaneously draw on more than one ability. Even though a specific ability may be the focus of the test, other abilities will contribute to the relation with false-belief understanding. For instance, children require semantic ability to comprehend the words used in syntax tests and thus, as Ruffman et al. (2003) argued, these tests are not a pure measure of syntactic ability. In contrast, receptive vocabulary tests simply require the child to point, for example, to one of four pictures, in order to indicate which one depicts a single word spoken by the experimenter. It may be that receptive vocabulary tests provide the purest measure of an isolated language ability because there is less overlap with other abilities, and this may at least partly explain why the effect size is the lowest for receptive vocabulary.

Type of False-belief Task

The relation between language ability and false-belief task performance ranges from $r = .35$ to $.52$ across the four different task types (i.e., change-of-

location, unexpected change, deception, and belief-emotion, see Table 6). However, even though the omnibus random effects test across the four types yielded a significant main effect of task type, the results of post hoc pair-wise comparisons indicated no significant differences between the types of false-belief task. That is, the effect sizes are not significantly different from one another. As described in the introduction, the false-belief tasks differ in a number of ways in terms of the linguistic demands of the tasks. Some involve a narrative whereas others do not, and some involve a test question with an embedded sentential complement whereas others use a simpler test question. Nonetheless, performance on the different types of task does not differ significantly by language ability. This result confirms the findings of individual studies that show no difference between the change-of-location and unexpected-identity false-belief tasks in terms of their relation to language (e.g., Jenkins & Astington, 1996). Given the variability in results found in the relation between language ability and false-belief task performance for the 104 studies included in the meta-analysis, the absence of a significant difference between false-belief task types when results of studies are combined provides confidence in this conclusion. This is because the combination of results in meta-analysis increases the power to detect patterns of effect over a number of studies. In addition, our data corroborate previous research that demonstrated a relation between language and belief-based emotion-attribution tasks (de Rosnay, Pons, Harris, & Morrell, 2004).

The finding of no significant difference for different false-belief task types, in their relation to language, is comparable to that in Wellman et al.'s (2001) meta-analysis showing no age difference in children's performance on the change-of-location and unexpected-identity false-belief tasks. Wellman et al. maintain that their finding supports the argument that there is a genuine conceptual change underlying performance on different types of false-belief task, that is, the development of a concept of belief and an understanding that beliefs represent and may misrepresent the world. Our data support their interpretation, and indeed strengthen it with the addition of two other types of task that were not included in the Wellman et al. meta-analysis (viz., deception and belief-emotion). Nonetheless, it is possible that we did not find a significant difference between task types because even within task type there is also a variety of linguistic demands, as described in the introduction.

The fact that there is a significant relation with language for each false-belief task type, despite their varying linguistic demands, provides some support

for the argument that the relation between language ability and false-belief understanding is not an artifact of task factors. Even so, at least part of the effect may well be due to the fact that all of the false-belief tasks used in studies included in the meta-analysis were verbal tasks, albeit making greater or lesser verbal demands on the child. The present findings would be greatly strengthened by demonstration of a relation between language ability and performance on nonverbal false-belief tasks. Such a relation has been demonstrated for deaf children using almost nonverbal (so-called "low verbal") false-belief tasks (P. de Villiers, 2005). Although comparable data from typically developing children are not reported in the literature, we would expect to find the same relation, given the argument that language supports the conceptual understanding of false belief. However, whether there would be a relation between language ability and implicit understanding of false belief is open to question. Clements and Perner (1994) used the standard change-in-location task procedure with children younger than those who typically pass false-belief tasks. They showed that children look toward the correct (i.e., the initial) location of the object in anticipation of the protagonist's search; yet they give an incorrect response to the standard verbal test question. In another modified false-belief task that uses violation-of-expectation methodology, Onishi and Baillargeon (2005) showed that infants look longer in conditions that violate the expectation that an actor's search is premised on the actor's beliefs, not on the actual situation. Such findings support the argument that children fail standard false-belief tasks because they lack the requisite linguistic and computational resources (Bloom & German, 2000; Fodor, 1992). It is still the case, however, that success on the standard tasks may mark genuine conceptual development of the sort Karmiloff-Smith (1992) calls "representational redescription," where knowledge that is at first implicit in procedures becomes explicit and available to conscious access. Moreover, language may be critical to this process (Astington, 2006), in which case we might expect to find weaker relations between language ability and performance on implicit false-belief tasks, alongside typical strong relations between language and performance on nonverbal versions of standard false-belief tasks. This research is still to be conducted.

Direction of Effect

Regarding the direction of effect in longitudinal studies, the results of the present meta-analysis show a significant effect size for both directions of relation.

That is, earlier performance on measures of language ability predicts later false-belief task performance, as well as the reverse—earlier false-belief task performance predicts later performance on language ability measures. This finding provides support for the argument that the relation between language ability and false-belief understanding is bidirectional (Slade & Ruffman, 2005). The direction of effect from language to false belief would be expected by researchers who argue that language plays a causal role in the development of false-belief understanding (Astington & Baird, 2005a). Then again, the opposite direction of effect, from false belief to language, might also be expected because false-belief understanding plays a role in language acquisition; for example, children consider beliefs in word learning tasks, that is, they take speakers' knowledge and ignorance into account (Sabbagh & Baldwin, 2001).

Nonetheless, it is important to note that we found a significant difference between the two effect sizes, that is, the direction of effect is stronger from language ability to false-belief understanding than the reverse. This finding corroborates results from some longitudinal studies (Astington & Jenkins, 1999; J. de Villiers & Pyers, 2002), which show that language ability is a significant predictor of change in false-belief task performance, whereas the reverse is not the case, that is, false-belief understanding does not predict changes in language task performance. This provides support for the argument that language plays a role in the development of false-belief understanding, such that language ability will predict false-belief task performance over time because it provides children with the resources required to represent and communicate about mistaken beliefs.

However, Slade and Ruffman (2005) point out that the direction of effect from language to false belief might be explained on statistical rather than theoretical grounds. That is to say, there is often, although not always, a greater range in the language measure scores than in the false-belief task scores, which makes it more likely that language would predict false belief over time, rather than the reverse. Yet in the current meta-analysis, effect sizes of the relation between language and false belief did not differ significantly by number of false-belief task trials used, which perhaps suggests that a greater range of false-belief task scores would not affect the magnitude of the relation between language and false belief.

Effect of Age

Age does not provide an explanation for variation in children's false-belief task performance in the

same way that language ability does. Age is not really an explanatory variable, but rather it is a proxy for various maturational factors that may explain variation, an important one of which is language ability. Age and language ability strongly covary in typically developing populations and much of the variance in false-belief understanding that is explained by language ability will be shared with age. It is therefore not surprising that the effect size for the relation between language ability and false-belief task performance is reduced in the secondary analysis using studies that controlled for age.

The significant variability among studies indicated in the overall analysis was not found when age was controlled for, despite a wide range of partial correlations ($-.14$ to $.63$). This may suggest that some of the variability in study effect sizes in the overall analysis is due to maturational factors. In the present meta-analysis, only 32 studies reported or provided data on the relation between language ability and false-belief understanding, controlling for age. It is possible that greater variability may have resulted if partial correlations from a larger number of studies were included, particularly given the wide range of partial correlations found in the present analysis. Nonetheless, both the effect size and the nonsignificant variability found highlight the crucial need to control for age when assessing the relation between language ability and false-belief understanding. Unfortunately, only a minority of studies do so.

Role of Language in Development of False-belief Understanding

The present meta-analysis focuses on the relation between language ability, measured using standardized inventories and experimental tasks, and performance on false-belief tasks for typically developing, English-speaking children. We restricted our investigation in this way for the reasons presented in the introduction, that is, in order to focus on the contribution of individual linguistic ability and to limit the number of variables included in the analyses. However, the limitations of our findings are thus evident. A meta-analysis of studies using a range of clinical populations, as well as a meta-analysis of studies conducted in languages other than English, would certainly add weight to our findings if they were replicated. We anticipate that they would be, given that significant relations between language measures and children's false-belief task performance have been found in studies with clinical samples, including children with autism,

deafness, specific language impairment, Asperger syndrome, and Down syndrome (e.g., P. de Villiers, 2005b; Happé, 1995; Miller, 2001; Peterson & Siegal, 1999; Tager-Flusberg & Joseph, 2005). Significant relations between language and false belief are also reported for studies conducted in numerous languages other than English, such as Chinese, French, German, Italian, Japanese, Spanish, Swedish, and Turkish (e.g., Antonietti, Liverta-Sempio, Marchetti, & Astington, 2006; Cheung et al., 2004; Deleau, Hooge, & Bernard, 2002; Perner, Lang, & Kloo, 2002; Yagmurlu, Berument, & Celimli, 2005).

Role of conversation. In restricting our analysis to standardized language inventories and experimental language tasks we highlight the contribution of the individual child's linguistic ability to the development of false-belief understanding. This is, however, only part of the whole picture. Pragmatic and semantic measures derived from children's conversations add an important new dimension. Pragmatic measures, which assess children's ability to use language in communicative exchanges, are for obvious reasons mostly derived from naturalistic observation of conversation. A valuable pragmatic measure is connectedness of communication, that is, the degree to which a child's speech acts are linked to the interlocutor's previous utterances (Dunn & Brophy, 2005). Successful connected communication predicts successful false-belief task performance both concurrently and longitudinally (Dunn & Cutting, 1999; Slomkowski & Dunn, 1996). This finding would be expected, given that connected communication depends on taking account, even if only implicitly, of the point of view of one's conversational partner. Indeed, generally, there is a close conceptual link between pragmatics and theory of mind because pragmatic ability depends on keeping track of speaker/listener beliefs and intentions in order to communicate successfully.

Production of mental terms is the semantic measure most frequently derived from conversation in order to assess the relation between language and false-belief understanding. Broadly speaking, this is the set of terms that refer to mental states of belief, desire, intention, emotion, and perception. Many different classification schemes are used in the literature, and the phrase "mental terms" may be used to refer just to cognitive terms (e.g., *think*, *know*, *guess*, *remember*). Children's use of cognitive terms, as well as mental terms more generally, has been shown to correlate with false-belief task performance both concurrently and across time, for example, in conversations with friends in the context of pretend play (Hughes & Dunn, 1998), and in conversations with

mothers while looking at photographs (Ruffman et al., 2002). Child and parent uses of mental terms are related, and much of the research on conversation has investigated the effect of parental language on children's development of false-belief understanding. For example, mothers' use of mental terms is related to their children's later success on false-belief tasks, controlling for the children's earlier false-belief task performance and language ability (Ruffman et al., 2002). Importantly, some mental terms, such as *think*, are used in conversation in sentential complement constructions. As described in the introduction, a sentential complement is a tensed subordinate clause that is embedded under a mental or communication verb to form a complex sentence, for example: "Maxi thinks that the chocolate is in the cupboard." This syntactic structure provides the linguistic format needed to represent false beliefs and therefore it has been of great interest in the debate over the role of language in the development of false-belief understanding.

Role of sentential complements. In the present meta-analysis the largest effect size ($r = .66$) was found for the relation between false belief and performance on the memory for complements test (J. de Villiers & Pyers, 2002), although this correlation did not differ significantly from that for other language measures. Some researchers (Cheung et al., 2004; Ruffman et al., 2003) argue that this particular correlation is not surprising because there is an overlap in the requirements for the false-belief test and the memory for complements test. The latter test presents children with stories about mistakes or lies that are described in complex sentences consisting of a sentential complement embedded under a mental main verb (e.g., *think*, *believe*) or a communicational main verb (e.g., *say*, *tell*). For example: "He thought [or said] he found his ring, but it was really a bottle cap. What did he think [or say]?" Ruffman et al. argue that children cannot understand such stories and respond appropriately unless they understand false belief. J. de Villiers (2005b) disputes this point, particularly for test items that use verbs of communication like *say*. However, even there, children may need to understand false belief in order to respond correctly to a question about what someone said, in cases where the child knows the person holds a false belief, as Wimmer and Hartl (1991, Experiment 3) showed. Indeed, Wellman et al.'s (2001) meta-analysis confirms this finding, showing no significant difference among false-belief test questions using mental or communication verbs.

In fact, there is much debate in the literature concerning the role of children's mastery of sentential

complements in the development of false-belief understanding (Astington & Baird, 2005a; Cheung et al., 2004; Perner, Sprung, Zauner, & Haider, 2003). Yet, remarkably, of the 104 studies included in the present meta-analysis, only four (two unpublished) assess children's understanding of sentential complements, and one of these (Cheung et al., 2004) uses a memory for complements test that does not actually address J. de Villiers's (2005) theory. Importantly, the theory does not apply to all sentential complements, but only to those embedded under mental and communication verbs taking *realis* complements (i.e., about real events). Thus, it does not apply to [*promise-that*] in English (Cheung et al., 2004) or to [*want-that*] in German (Perner et al., 2003) because these verbs take *irrealis* complements (i.e., about unrealized—future or hypothetical—events). Critically, *realis* complements are true or false, whereas *irrealis* complements cannot be either true or false. The key point in J. de Villiers's theory is that children first master sentential complements embedded under communication verbs (e.g., *say*), where there can be overt evidence that the complement is false, and then by analogy they come to understand false complements embedded under mental verbs (e.g., *think*). The theory is not just about the syntax of sentential complements, but rather it involves semantics and syntax of mental and communication verbs. It is, indeed, "... a very specific hypothesis about the emergence of false-belief understanding, namely, that it rests on the child's mastery of the grammar (*semantics and syntax*) of complementation" (J. de Villiers & Pyers, 2002, p. 1040, emphasis added). Strong evidence in favor of J. de Villiers's (2005) theory would be a training study that demonstrated that children who were exposed to indirect quotation of statements they know to be false (e.g., "Dad says it's raining" when they know it is sunny; i.e., sentential complement constructions using *say*) improved their performance on nonverbal versions of standard false-belief tasks, in comparison with an appropriate control group. In fact, more generally, training studies might provide answers to some of the outstanding questions about the role of language in the development of false-belief understanding.

Training studies. The present meta-analysis investigates both the magnitude of the overall correlation between language and false-belief understanding and the factors that affect variability among individual correlations. A limitation of the study is that we were not able to perform fine-grained analyses to investigate relations among the variables considered in the secondary analyses, that is, types of language measure, types of false-belief task, and directions of

effect. This is the kind of detailed systematic analysis that is now required. But even so, investigation of correlations, even cross-lagged correlations in longitudinal studies, cannot provide conclusive evidence of causation. Causal explanation depends on a thoughtful combination of correlational data and information from training studies (Bryant, 1990).

There are already a number of training studies in the literature that address the role of language in the development of false-belief understanding. For example, children's false-belief task performance improves, relative to untrained groups, if they are engaged in conversations about characters in story books or videos, who play tricks to surprise or deceive one another (Appleton & Reddy, 1996; Guajardo & Watson, 2002). However, in these conversations children's attention is focused on the characters' different perspectives but using mental verbs with sentential complements. It is unclear whether perspective-taking, or mental-verb semantics, or complement syntax is the driving force behind the training effect. More recent studies have attempted to disentangle these factors. Children trained using false complement constructions with the verb *say* showed improved performance on false-belief tasks relative to a control group (Hale & Tager-Flusberg, 2003). This finding supports J. de Villiers's (2005) theory by showing that overt evidence of falsity in statements leads to an understanding of falsity in belief. Furthermore, in a study conducted in German (Lohmann & Tomasello, 2003), children's false-belief task performance improved in three training conditions: conversation about deceptive objects (e.g., a pen that looked like a flower) using no sentential complements; training on the syntax of complementation without the deceptive objects; and third, where there was the largest effect, using sentential complements with *think* and *say* in conversation about the deceptive objects. A fourth condition, where children were shown the deceptive objects without conversation, had no effect. These findings suggest that dyadic conversation and individual mastery of complementation syntax make independent contributions to the development of false-belief understanding, and furthermore, they show that language is needed because the perceptual evidence that appearances can be deceiving was not sufficient.

Together, these findings make an important contribution to the debate on the role of language in the development of false-belief understanding. However, additional training studies need to be designed and conducted to show that the reported effects occur in other populations (e.g., other language

groups) and to address the outstanding questions. For example, it is important to discover whether language training would lead to improved performance on nonverbal false-belief tasks and, if so, which particular aspects of language are most effective. For instance, training studies could further separate the semantic and syntactic elements of sentential complement training in order to discover whether the effect is primarily due to falsity or to embedded syntax. Training studies could also help to further separate the contributions of communicative social interaction and individual language development. It is likely that both of these two factors play a role in the development of false-belief understanding and, indeed, there may be individual differences in their relative importance. This is suggested by Jenkins and Astington's (1996) finding that the false-belief test performance of children with lower language ability and two siblings was equal to that of children with higher language ability and no siblings, whereas children with lower language abilities and no siblings performed more poorly. This implies that individual language skills and sibling social interaction both contribute to the development of false-belief understanding and suggests that either one of these can compensate for a deficit in the other.

Altogether, the additional meta-analyses and training studies that we have proposed would help to elucidate why language matters for the development of false-belief understanding, in particular, and for theory-of-mind development more generally considered. The ability to interpret people's behavior on the basis of their desires and emotions often depends on taking account of their beliefs as well. This is seen, for example, in the belief-emotion task included in the present meta-analysis, where false belief has to be considered in order to make a correct emotion prediction based on desire.

The results of the present meta-analysis show that there is a strong relation between false-belief understanding and language ability, which holds across a variety of language ability measures and false-belief task types, both concurrently and longitudinally, with a stronger direction of effect from language to false-belief than the reverse. These findings provide support for the argument that language plays a vital role in the development of false-belief understanding, and thus in the development of theory of mind.

References

Note. References preceded by an asterisk indicate studies included in the meta-analysis.

- *Ain (Dack), L. (2004). *Young children's deontic and epistemic reasoning*. Unpublished master's thesis, University of Toronto, Toronto, Ontario, Canada.
- *Angelopoulos, M., & Moore, C. (2003). *Social reputation in preschoolers and its relation to theory of mind, sharing, and prudence*. Unpublished manuscript, Dalhousie University, Halifax, NS, Canada.
- Antonietti, A., Liverta-Sempio, O., Marchetti, A., & Astington, J. W. (2006). Mental language and understanding of epistemic and emotional mental states: Contextual aspects. In A. Antonietti, O. Liverta-Sempio, & A. Marchetti (Eds.), *Theory of mind and language in developmental contexts* (pp. 1–30). Springer: Dordrecht, the Netherlands.
- Appleton, M., & Reddy, V. (1996). Teaching 3-year-olds to pass false-belief tests: A conversational approach. *Social Development*, 5, 275–291.
- *Artuso, M. (2000). Information processing constraints on the acquisition of a theory of mind. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 60, 3601.
- *Astington, J. W. (2000). Language and metalanguage in children's understanding of mind. In J. W. Astington (Ed.), *Minds in the making: Essays in honor of David R. Olson* (pp. 267–284). Oxford, UK: Blackwell.
- Astington, J. W. (2006). The developmental interdependence of theory of mind and language. In S. C. Levinson & N. J. Enfield (Eds.), *The roots of human sociality: Culture, cognition, and human interaction* (pp. 179–206). Oxford, UK: Berg.
- Astington, J. W., & Baird, J. A. (2005a). Introduction: Why language matters. In J. W. Astington & J. Baird (Eds.), *Why language matters for theory of mind* (pp. 3–25). New York: Oxford University Press.
- *Astington, J. W., & Baird, J. A. (2005b). Representational development and false-belief understanding. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 163–185). New York: Oxford University Press.
- *Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory of mind development. *Developmental Psychology*, 35, 1311–1320.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.
- *Benigno, J. P. (2004). *Preschoolers as social facilitators: The relationship between preschoolers' theory of mind and the quality of interactions they share with younger siblings*. Unpublished doctoral dissertation, University of Florida.
- *Bialystock, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development*, 75, 562–579.
- *Blair, C., Granger, D., & Razza, R. P. (2005). Cortisol reactivity is positively related to executive function in preschool children attending Head Start. *Child Development*, 76, 554–567.

- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false-belief task as a test of theory of mind. *Cognition*, 77, 25–31.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis, Version 2*. Englewood, NJ: Biostat.
- *Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the mind or predicting signal-dependent action? Performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development*, 6, 259–283.
- *Browning, C. A., & Holmes-Lonergan, H. A. (1998). *Relationships among semantic and syntactic language abilities and false-belief performance in young children*. Manuscript submitted for publication.
- Bryant, P. (1990). Empirical evidence for causes in development. In G. Butterworth & P. Bryant (Eds.), *Causes of development* (pp. 33–45). London: Harvester Wheatsheaf.
- *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.
- Carlson, S. M., & Moses, L. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72, 1032–1053.
- *Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, 87, 299–319.
- *Cassidy, K. W., Fineberg, D. S., Brown, K., & Perkins, A. (2005). Theory of mind may be contagious, but you don't catch it from your twin. *Child Development*, 76, 97–106.
- *Cassidy, K. W., Werner, R. S., Rourke, M., & Zuber, L. S. (2003). The relationship between psychological understanding and positive social behaviors. *Social Development*, 12, 198–221.
- *Chan, K. T. (2004). *Chinese-English bilinguals' theory-of-mind development*. Unpublished doctoral dissertation, University of Toronto, Canada.
- Chandler, M. J., Fritz, A. S., & Hala, S. M. (1989). Small scale deceit: Deception as a marker of 2-, 3- and 4-year-olds' early theories of mind. *Child Development*, 60, 1263–1277.
- *Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false-belief development? *Social Development*, 11, 1–10.
- *Cheung, C. (2006). *English- and Cantonese-speaking children's understanding of emotions and false belief*. Unpublished master's thesis. University of Toronto, Canada.
- *Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (2004). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development*, 75, 1155–1170.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9, 377–395.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Lawrence Erlbaum.
- *Curenton, S. M. (2003). Low-income preschoolers' false-belief performance. *Journal of Genetic Psychology*, 164, 411–424.
- *de Rosnay, M., & Harris, P. L. (2002). Individual differences in children's understanding of emotion: The roles of attachment and language. *Attachment and Human Development*, 4, 39–54.
- *de Rosnay, M., Pons, F., Harris, P. L., & 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.
- de Villiers, J. G. (2005). Can language acquisition give children a point of view? In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 186–219). New York: Oxford University Press.
- *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.
- de Villiers, P. (2005). The role of language in theory-of-mind development: What deaf children tell us. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 266–297). New York: Oxford University Press.
- *Deák, G. O., & Enright, B. (2006). Choose and choose again: Appearance-reality errors and the logic of questioning. *Developmental Science*, 9, 323–333.
- *Deák, G. O., Ray, S. D., & Brenneman, K. (2003). Children's preservative appearance-reality errors are related to emerging language skills. *Child Development*, 74, 944–964.
- Deleau, M., Hooge, F., & Bernard, S. (2002, August). *Conversational awareness and representation of beliefs in young children*. Paper presented at the Meeting of the International Society for the Study of Behavioural Development, Ottawa, Canada.
- *Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, 27, 367–392.
- Dunn, J., & Brophy, M. (2005). Communication, relationships, and individual differences in children's understanding of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 50–69). New York: Oxford University Press.
- Dunn, J., & Cutting, A. L. (1999). Understanding others, and individual differences in friendship interactions in young children. *Social Development*, 8, 201–219.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test—3rd ed. (PPVT-III)*. Circle Pines, MN: American Guidance Service.

- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale—2nd ed. (BPVS-II)*. London: NFER-Nelson.
- *Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Journal of Applied Developmental Psychology, 24*, 51–73.
- *Farrar, M. J., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language, 22*, 197–213.
- Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist, 41*, 418–425.
- *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., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false-belief understanding and inhibition skills. *Developmental Science, 7*, 103–115.
- Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44*, 283–296.
- *Fonagy, P., Redfern, S., & Charman, T. (1997). The relationship between belief-desire reasoning and a projective measure of attachment security (SAT). *British Journal of Developmental Psychology, 15*, 51–61.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–528.
- *Gal Endres, L. M. (2003). *Role of theory of mind in the social competence of preschoolers*. Unpublished doctoral dissertation, Miami University.
- *Garner, P. W., Curenton, S. M., & Taylor, K. (2005). Predictors of mental state understanding in preschoolers of varying socioeconomic backgrounds. *International Journal of Behavioral Development, 29*, 271–281.
- *Garnham, W. A., Brooks, J., Garnham, A., & Ostenfeld, A. M. (2000). From synonyms to homonyms: Exploring the role of metarepresentation in language understanding. *Developmental Science, 3*, 428–441.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*, 26–37.
- *Greig, A., & Howe, D. (2001). Social understanding, attachment security of preschool children and maternal mental health. *British Journal of Developmental Psychology, 19*, 381–393.
- *Guajardo, N. R., & Turley-Ames, K. J. (2003). *Preschoolers' generation of different types of counterfactual statements and theory of mind understanding*. Manuscript submitted for publication. Published as: Guajardo & Turley-Ames, (2004).
- Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development, 19*, 53–80.
- Guajardo, N. R., & Watson, A. C. (2002). Narrative discourse and theory of mind development. *Journal of Genetic Psychology, 163*, 305–325.
- *Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Journal of Cognition and Development, 4*, 275–298.
- Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*, 346–359.
- Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*, 843–855.
- Harris, P. L. (2005). Conversation, pretense, and theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 70–83). Oxford University Press.
- Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition and Emotion, 3*, 379–400.
- *Hendry, C.-A. (2005). *Parental use of mental state language in shared book reading*. Unpublished master's thesis. University of Guelph, Canada.
- *Henseler, S. E. (2000). Young children's developing theory of mind: Person reference, psychological understanding and narrative skill. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*, 1110.
- *Holmes, A. M. (2002). Theory of mind and behaviour disorders in children with specific language impairment. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*, 5376.
- Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *Test of early language development: 3rd ed. (TELD-III)*. Austin, TX: Pro-Ed.
- *Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.
- *Hughes, C., & Cutting, A. L. (1999). Nature, nurture, and individual differences in early understanding of mind. *Psychological Science, 10*, 429–432.
- Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34*, 1026–1037.
- *Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. *Journal of Child Psychology and Psychiatry, 39*, 981–994.
- *Hughes, C., Jaffee, S. R., Happe, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2003). *Origins of individual differences in theory of mind: From nature to nurture?* Manuscript submitted for publication. Published as: Hughes, C., Jaffee, S. R., Happe, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development, 76*, 356–370.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for

- cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275–292.
- *Jackson, A. L. (2001). Language facility and theory of mind development in deaf children. *Journal of Deaf Studies and Deaf Education*, 6, 161–176.
- *James, T. L. (2002). Relations between pragmatic development and theory of mind in young children. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 63, 2130.
- *Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology*, 32, 70–78.
- *Joe, W. Q. (2003). *An executive performance account of theory of mind expression*. Unpublished doctoral dissertation. Bryn Mawr College.
- *Kamawar, D. (2000). Children's understanding of the opaque and transparent uses of language. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 61, 3307.
- *Kamawar, D., & de Villiers, J. (2002). [Opaque contexts: Factors contributing to success.] Unpublished raw data, Carleton University, Ottawa, Canada.
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.
- *Keenan, T. (1998). Memory span as a predictor of false-belief understanding. *New Zealand Journal of Psychology*, 27, 36–43.
- *Liebman, J. I. (2005). The developmental patterns and cognitive predictors of prosocial display rule behavior and understanding: Theory of mind and inhibitory control. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 66, 1759.
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false-belief understanding: A training study. *Child Development*, 74, 1130–1144.
- *Maring, B. L. (2003). *The metaphorical bases of children's developing theories of mind*. Unpublished doctoral dissertation. University of Oregon.
- *Mathews, R., Dissanayake, C., & Pratt, C. (2003). The relationship between theory of mind and conversation abilities in children using an active/inactive paradigm. *Australian Journal of Psychology*, 55, 35–42.
- *McNab, C. (2001). *Peer acceptance and theory of mind*. Unpublished master's thesis, University of Canterbury.
- *Meins, E., Fernyhough, C., Russell, J., & Clark-Carter, D. (1998). Security of attachment as a predictor of symbolic and metalising abilities: A longitudinal study. *Social Development*, 7, 1–24.
- *Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development*, 73, 1715–1726.
- *Miller, C. A. (2001). False-belief understanding in children with specific language impairment. *Journal of Communication Disorders*, 34, 73–86.
- *Müller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development*, 20, 173–189.
- *Mullins, D. (2004). *A comparison of false belief and referential opacity tasks: Sorting out the relations*. Unpublished master's thesis, Carleton University, Canada.
- *Murray, L., Woolgar, M., Briers, S., & Hipwell, A. (1999). Children's social representations in dolls' house play and theory of mind tasks, and their relation to family adversity and child disturbance. *Social Development*, 8, 179–200.
- *Nash, H. M. (2001). *Social skills training, training to task, and theory of mind performance among young children*. Unpublished doctoral dissertation, Idaho State University.
- Nelson, K. (2005). Language pathways into the community of minds. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 26–49). New York: Oxford University Press.
- *Nelson, K., Plesa, D., Goldman, S., Henseler, S., Presler, N., & Walkenfeld, F. F. (2003). Entering a community of minds: An experiential approach to "theory of mind". *Human Development*, 46, 24–46.
- Onishi, K. H., & Baillargeon, R. (2005). 15-month-old infants understand false beliefs. *Science*, 308, 255–257.
- *Pears, K. C., & Fisher, P. A. (2005). Emotion understanding and theory of mind among maltreated children in foster care: Evidence of deficits. *Development and Psychopathology*, 17, 47–65.
- *Pears, K. C., & Moses, L. J. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development*, 12, 1–20.
- *Pellicano, E., Maybery, M., & Durkin, K. (2005). Central coherence in typically developing preschoolers: Does it cohere and does it relate to mindreading and executive control? *Journal of Child Psychology and Psychiatry*, 46, 533–547.
- Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self control: More than a common problem of inhibition. *Child Development*, 73, 752–767.
- Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5, 125–137.
- Perner, J., Sprung, M., Zauner, P., & Haider, H. (2003). *Want that* is understood well before *say that*, *think that*, and false belief: A test of de Villiers' linguistic determinism on German-speaking children. *Child Development*, 74, 179–188.
- *Peterson, C. C. (2000). Kindred spirits: Influences of siblings' perspectives on theory of mind. *Cognitive Development*, 15, 435–455.
- *Peterson, C. C. (2004). Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry*, 45, 1096–1106.

- Peterson, C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf, and normal-hearing children. *Psychological Science*, 10, 126–129.
- *Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, 18, 399–429.
- *Razza, R. A. (2005). Relations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 66, 2859.
- *Renouf, A., Brendgen, M., Seguin, J. R., Boivin, M., Dionne, G., Vitaro, F., et al. (2006). [Empathy as a moderator of the relation between theory of mind and indirect aggression.] Unpublished raw data. University of Montreal, Montreal, Canada.
- Repacholi, B., & Slaughter, V. (Eds.) (2003). *Individual differences in theory of mind: Implications for typical and atypical development*. New York: Psychology Press.
- *Repacholi, B., & Trapolini, T. (2004). Attachment and preschool children's understanding of maternal versus non-maternal psychological states. *British Journal of Developmental Psychology*, 22, 395–415.
- Rosenthal, M. C. (1994). The fugitive literature. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 85–94). New York: Russell Sage Foundation.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- *Ruffman, T., Perner, J., & Parkin, L. (1999). How parenting style affects false-belief understanding. *Social Development*, 8, 395–411.
- *Ruffman, T., Slade, L., Clements, W., & Import, A. (1998). *How language is related to verbal and nonverbal theory of mind*. Unpublished manuscript, University of Sussex.
- *Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, 73, 734–751.
- *Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2001) *How language relates to belief, desire, and emotion understanding*. Unpublished manuscript. Published as: Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development*, 18, 139–158.
- Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development*, 18, 139–158.
- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, 72, 1054–1070.
- *Schwebel, D. C., Rosen, C. S., & Singer, J. L. (1999). Preschoolers' pretend play and theory of mind: The role of jointly constructed pretence. *British Journal of Developmental Psychology*, 17, 333–348.
- *Seamans, E. L. (2004). *Parent-child conversations and the development of representational theory of mind*. Unpublished master's thesis, Queen's University, Canada.
- Shatz, M. (1994). Theory of mind and the development of social linguistic intelligence in early childhood. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind* (pp. 311–329). Hove, UK: Erlbaum.
- Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, 14, 301–321.
- Slade, L., & Ruffman, T. (2003) *How language does (and does not) relate to theory-of-mind: A longitudinal study of syntax, semantics, working memory and false belief*. Unpublished manuscript. Published as: Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory-of-mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, 23, 117–141.
- Slomkowski, C., & Dunn, J. (1996). Young children's understanding of other people's beliefs and feelings and their connected communication with friends. *Developmental Psychology*, 32, 442–447.
- *Smith, M., Apperly, I., & White, V. (2003). False-belief reasoning and the acquisition of relative clause sentences. *Child Development*, 74, 1709–1719.
- *Symons, D. K., Peterson, C. C., Slaughter, V., Roche, J., & Doyle, E. (2003) *Theory of mind and mental state discourse during book reading and story-telling tasks*. Unpublished manuscript. Published as: Symons, D. K., Peterson, C. C., Slaughter, V., Roche, J., & Doyle, E. (2005). Theory of mind and mental state discourse during book reading and story-telling tasks. *British Journal of Developmental Psychology*, 23, 1–24.
- Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–318). New York: Oxford University Press.
- *Taylor, M., & Carlson, S. M. (1997). The relation between individual differences in fantasy and theory of mind. *Child Development*, 68, 436–455.
- Varley, R. (1998). Aphasic language, aphasic thought: An investigation of propositional thinking in an a-propositional aphasic. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 128–145). Cambridge, UK: Cambridge University Press.
- Varley, R., Siegal, M., & Want, S. C. (2001). Severe impairment in grammar does not preclude theory of mind. *Neurocase*, 7, 489–493.
- *Watson, A. C., Nixon, C. L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental Psychology*, 35, 386–391.
- *Watson, A. C., Painter, K. M., & Bornstein, M. H. (2001). Longitudinal relations between 2-year-olds' language and 4-year-olds' theory of mind. *Journal of Cognition and Development*, 2, 449–457.

- *Watson, J. K. (2000). Theory of mind and pretend play in family context. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, 60, 3599.
- *Weimer, A. A., & Guajardo, N. R. (2005). False-belief, emotion understanding, and social skills among Head Start and non-Head Start children. *Early Education and Development*, 16, 341–366.
- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167–187). Oxford, UK: Blackwell.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655–684.
- *Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, 7, 283–288.
- *Whitehouse, A. J. O., & Hird, K. (2004). Is grammatical competence a precondition for belief-desire reasoning? Evidence from typically developing children and those with autism. *Advances in Speech-Language Pathology*, 6, 39–51.
- Wiig, E. H., Secord, W., & Semel, E. (1992). *Clinical evaluation of language fundamentals—preschool (CELF—Preschool)*. San Antonio, TX: The Psychological Corporation, Harcourt Brace Jovanovich.
- *Wilson, A. (1998). *Theory of mind development and attentional processes in young school children*. Unpublished thesis, West Virginia University.
- Wimmer, H., & Hartl, M. (1991). Against the Cartesian view on mind: Young children's difficulty with own false belief. *British Journal of Developmental Psychology*, 9, 125–138.
- 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.
- Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development*, 73, 768–778.
- Yagmurlu, B., Berument, S. K., & Celimli, S. (2005). The role of institution and home contexts in theory of mind development. *Journal of Applied Developmental Psychology*, 26, 521–537.
- *Ziadas, K., Durkin, K., & Pratt, C. (1998). Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: Links to theory of mind development. *Journal of Child Psychology and Psychiatry*, 39, 755–763.