Meta-Analyses Comparing Theory of Mind Abilities of Individuals With Autism, Individuals With Mental Retardation, and Normally Developing Individuals

Nurit Yirmiya, Osnat Erel, Michal Shaked, and Daphna Solomonica-Levi
The Hebrew University of Jerusalem

A deficit in theory of mind (ToM) abilities has been described as the core deficit in autism. The authors performed 3 meta-analyses, comparing ToM abilities of individuals with autism, individuals with mental retardation (MR), and normally developing individuals. Results indicated that individuals with autism and MR have impaired ToM abilities. The etiology associated with MR (i.e., Down syndrome, undifferentiated etiology) was found to be an important moderator variable. Chronological age (CA) and verbal mental age (VMA) of the normally developing children and CA, VMA, and performance mental age of individuals with MR, and type of matching between the groups were also found to be moderator variables. Discussion focuses on the implication of the findings and emphasizes the need to consider the specific etiology of comparison groups when studying abilities and impairments of individuals with autism and MR.

In the past five decades since Kanner (1943) first described the syndrome of autism, researchers have been trying to describe and conceptualize the syndrome and to identify its underlying causes. A number of hypotheses were offered and tested experimentally in an effort to account for the unique deficit or deficits in autism. The theory of mind (ToM) hypothesis, a current and leading hypothesis in the field, posits that individuals with autism are impaired in their ability to understand mental states. Our aim in this article is to summarize the empirical literature on the ToM hypothesis.

For a deficit in ToM to be considered as primary or central to autism, it has to be universal (i.e., manifested in all or almost all individuals with autism), and it has to be unique (i.e., not manifested by most individuals with other clinical diagnoses; Sigman, 1994, 1996). In the current study, we applied three meta-analyses to examine whether the deficit in ToM is unique to autism as compared to ToM abilities of individuals with mental retardation (MR) and those of normally developing children.

ToM refers to the ability to attribute mental states to the self and to others. The ability to know about minds is required for all human interactions; it is necessary for understanding, explaining, predicting, and manipulating the behavior of others (for comprehensive reviews and collections, see Astington, Har-

ris, & Olson, 1988; Baron-Cohen, Tager-Flusberg, & Cohen, 1993; Butterworth, Harris, Leslie, & Wellman, 1991; Frye & Moore, 1991; Lewis & Mitchell, 1994; Perner, 1991; Taylor, 1996; Wellman, 1990; Whiten, 1991). According to Wellman (1993), this ability involves two components: the ontological aspect, that is, the ability to distinguish between the real and the mental world, and the causal aspect, that is, the ability to understand mutual causal relations between mental states and the physical behavioral world. The acquisition of ToM abilities in normal children is one of the major developmental achievements of the first few years of life.

In trying to trace the development of ToM abilities in young children, most researchers have focused on the period between $2\frac{1}{2}$ to 5 years of age. Using various paradigms, these investigators suggest that the development of ToM abilities in normal children follows a typical pattern of gradual changes. Signs of this ability are apparent by 18 months, when children engage in symbolic play. This type of play involves behaviorally treating an object as symbolizing something other than that which it is known to be (Harris, 1991; Lillard, 1993, 1996; Moore, 1996) or mentally representing it by decoupling it from its referent (Leslie, 1988). By the age of 3 years, children begin to acquire a broader understanding of mental states: They are better able to understand the difference between their own mental states and those of others; they seem to understand the limits of what others can perceive; and they understand that certain perceptions restrict thoughts, and thus that people may differ in what they see, know, expect, like, and want (Harris, Taylor, 1996). Yet, 3-year-olds still make a firm differentiation between two types of mental states. On one hand, there are mental states such as perceptions, desires, and beliefs, which are expected to be strongly related to reality. On the other hand, there are mental states such as images, dreams, and pretenses, which the young child considers "silly" because they have no referential or causal relation to reality. Whereas most 4-year-olds are aware that this dichotomy does not always hold, the great majority of

3-year-olds consider perceptions, desires, and beliefs "transparent" in that they mentally reflect reality. This misconception of 3-year-olds is manifested by their difficulty with tasks demanding the understanding of false beliefs or the distinction between appearance and reality (Gopnik, 1993).

A general shift in the child's concept of the mind occurs between the ages of 3 and 4 years. During this period most children are able to understand false beliefs, begin to realize the distinction between appearance and reality, understand the concepts of desire and intention, and are aware of different sources of beliefs. At this age, children also begin to understand that people's actions are guided by their thoughts, beliefs, and desires and thus make use of the causal aspect of ToM (Flavell, 1992; Gopnik & Astington, 1988; Gopnik & Slaughter, 1991; Taylor, 1996; Wellman, 1993; Wimmer & Perner, 1983).

Several questions have been posed in the literature regarding the development of ToM in normal children. One question relates to the origin of ToM, that is, how much of ToM is built into the organism versus acquired through experience. A related question is how ToM is acquired and the specification of the early developmental precursors required for ToM (for a review, see Moore, 1996). Another question relates to the sequencing of acquiring ToM. There is some debate as to the exact age at which children acquire ToM abilities. Some researchers maintaining the early-onset view have suggested that by the age of 2 ½ years, children are able to use deceptive strategies, and that by age 3 years they are able to know about desires, values, and preferences and to answer specific and straightforward questions regarding ToM (Astington & Gopnik, 1991; Chandler, Fritz, & Hala, 1989; Flavell, 1992; Flavell, Flavell, Green, & Moses, 1990; Hala, Chandler, & Fritz, 1991; Lewis & Osbourne, 1990; Taylor, 1996). Other researchers, holding the late-onset view, have argued that true ToM abilities are present only after age 4 years (Ruffman, Olson, Ash, & Keenan, 1993; Russell, Mauthner, Sharpe, & Tidswell, 1991). It may be that some of the differences between the early- and late-onset views are associated with the tasks used to examine ToM abilities. Some tasks may be easier than others and can be passed at a younger chronological age (CA) or mental age (MA), whereas others are more difficult and can be passed only at an older CA or MA. Not only are some experimental tasks more difficult than others, but some researchers prefer naturalistic and behavioral observations over the more strict laboratory tasks, which are predominately based on verbal responses to carefully asked questions. Thus, there is ongoing debate as to the appropriate method or methods of data collection regarding the presence of ToM. Researchers also have debated whether children really acquire a theory or just a collection of concepts and rules for predicting and explaining behavior and thought (Flavell, 1992; Moore, 1996; Moses & Chandler, 1992; Taylor, 1996).

ToM in Autism

The syndrome of autism is manifested in a variety of symptoms forming a specific triad of impairments in communication, imagination, and socialization (American Psychiatric Association, 1994; Wing & Gould, 1979). Proponents of the ToM hypothesis of autism have claimed that one of the primary deficits in autism is a deficiency in the ability to attribute mental states

to oneself and others (Baron-Cohen, 1989b; Baron-Cohen, Leslie, & Frith, 1985; U. Frith, 1989; Hobson, 1993, 1990a, 1990b; Hughes & Russell, 1993; Leslie & Frith, 1988; Ozonoff, Pennington, & Rogers, 1991; Russell et al., 1991). Within this theoretical framework, the communication impairment can be understood as a problem in the semantics of mental states, the imagination impairment as a problem in attributing mental states that are contrary to reality, and the social impairment as an inability to understand the way in which mental states affect behavior. As a result, individuals with autism do not understand social situations and interact inappropriately (U. Frith; Leslie, 1987).

Children with autism, like normally developing 3-year-olds, fail on various tasks assessing the understanding of mental states. According to the ToM hypothesis, the failure of individuals with autism does not imply that they are simply operating like normally developing 3-year-olds (Baron-Cohen, 1992; Burack, 1992; Happé, 1994a, 1994c; Leslie & Roth, 1993; Roth & Leslie, 1991). Rather, because most individuals with autism do not perform at their expected MA, even long after they reach a MA of 4 years, a unique developmental deviance in ToM acquisition in autism has been suggested. This developmental deviance in ToM may have its earlier roots in the faulty development of earlier perceptual, cognitive, and social–emotional modules that are required for the development of ToM as suggested by researchers adhering to modularity theories (Baron-Cohen, 1994, 1995; Baron-Cohen & Ring, 1994; Leslie, 1994; Leslie & Thaiss, 1992).

Researchers testing ToM abilities in individuals with autism have frequently examined the understanding of false belief, using the Wimmer and Perner (1983) paradigm of Maxi or a variation of it, such as the Sally and Anne paradigm. In these experiments, participants are required to predict a protagonist's action on the basis of his or her absence during an important change in reality. For example, one doll named Sally leaves her ball in the red cupboard and leaves the scene. While she is away playing outside, the other doll, named Anne, moves the ball to the green cupboard. On Sally's return, participants are asked to predict where she will look for her ball. To answer correctly, the participant must understand that other people may hold beliefs that are both different from his or her own and that these beliefs are contrary to reality. The participant also must realize that these mental states determine people's actions.

A second common false-belief paradigm is the Smarties or milk-carton task, in which the participant is presented with contents different from what experience may lead one to expect (a Smarties box with a pencil rather than candy inside). The participant is then required either to recall his or her own past belief before being presented with the unexpected content or to predict another person's belief, again on the basis of this person's absence when the content was first presented. About 80% of the participants with autism do not attribute a false belief to the agent (e.g., Baron-Cohen 1989a; Baron-Cohen et al., 1985; Leslie & Frith, 1988; Perner, Frith, Leslie, & Leekam, 1989), whereas most normally developing 4-year-old children pass these tasks. Furthermore, the 20% of the participants with autism who pass the task do not pass tasks requiring second-order mental attributions, in which they are asked to predict the belief of one agent concerning another agent's belief (Baron-Cohen,

1989b; Bowler, 1992; Leekam & Prior, 1994; Ozonoff et al., 1991; Tager-Flusberg & Sullivan, 1994b).

A third paradigm commonly used in studying ToM abilities involves deception. The rationale underlying this paradigm is that deception involves manipulating others' thoughts or beliefs (Baron-Cohen, 1992; Hughes & Russell, 1993; Russell et al., 1991; Sodian, 1991; Sodian & Frith, 1992; Yirmiya, Solomonica-Levi, & Shulman, 1996; Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996). Additional paradigms include understanding various picture stories in which mental states are a central component (Baron-Cohen, Leslie, & Frith, 1986; Happé, 1994b; Tager-Flusberg & Sullivan, 1994b) and understanding mental–physical distinctions and the brain's function (Baron-Cohen, 1989b). In general, the findings reported in these studies are similar to those found in the false-belief tasks, with most participants with autism consistently failing.

Happé (1995) was the first to conduct a large-scale empirical summary of ToM abilities of 70 individuals with autism, 34 individuals with MR, and 70 normally developing children on two classical false-belief tasks: the Sally and Anne task and the Smarties task. Her analyses revealed that the percentage of participants with autism (20%) who passed the two ToM tasks was significantly lower than the percentage of participants who passed the two tasks in the normal (56%) and MR (58%) groups. The percentages of those who passed in the two nonautistic groups were not significantly different. For the normally developing group there was a 50% probability of passing the two tasks at a verbal mental age (VMA) of 4 years. For participants with autism, the same 50% probability of passing was reached only at a VMA of 9 years and 2 months. Happé suggested a two-threshold model in which there is a lower bound of VMA for both the normally developing individuals and for the individuals with autism below which none of the participants could pass both ToM tasks (a VMA of 2 years and 10 months for the normal group and 5 years and 6 months for the group with autism), and an upper bound of VMA above which all participants pass both tasks (a VMA of 6 years and 9 months for the normal group and 11 years and 7 months for the group with autism). This model was not applicable to the group with MR because no relation was found between ToM task performance and VMA in this group.

Happé's (1995) study was the first attempt to comprehensively summarize the ToM abilities of participants with autism, MR, and normal development. Happé's work, although addressing the association between ToM abilities and CA and VMA (as assessed by a one-word receptive language ability), does not address additional variables that may be associated with ToM abilities of individuals with autism and individuals with MR. For example, the specific diagnosis of individuals with MR may be important. These diagnoses were not reported by Happé, who wrote "The mentally handicapped subjects in this sample present something of a puzzle" (p. 851). The puzzle may have been solved if the specific etiologies and diagnoses of the individuals with MR were taken into account. Additional variables may also be important for studying ToM abilities in individuals with autism. For example, Happé examined performance on two ToM tasks, the Sally and Anne task and the Smarties task, and reported no significant group differences between the individuals with MR and the normally developing

participants. Yet other researchers reported significant differences in the ToM abilities of individuals with MR and normally developing children (e.g., Benson, Abbeduto, Short, Bibler-Nuccio, & Maas, 1993; Yirmiya, Solomonica-Levi, & Shulman, 1996; Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996; Zelazo, Burack, Benedetto, & Frye, 1996). It is also possible that some tasks are easier or more difficult for some but not all groups of individuals. Thus, type of task may be a relevant variable in understanding ToM abilities. In addition, Happé examined the association between VMA and ToM abilities. Other ability measures, such as performance MA, may be related differently among the various groups of individuals with autism and MR and may affect the magnitude of group differences if found.

Current Study

Our aim in the current study was twofold: The first goal was to examine whether the deficit in ToM is unique to autism in the context of comparisons with individuals with MR and normal development. Therefore, all published data in journals and dissertations on ToM abilities of individuals with autism, as compared to individuals with MR, and to normally developing children are considered, as well as all data on ToM abilities of individuals with MR as compared to normally developing children. On the basis of the literature, our hypotheses are that individuals with autism do not perform as well as normally developing children or individuals with MR grouped together regardless of etiology.

Given some recent reports that individuals with MR with unknown (nonspecific) etiology also show a deficit in ToM abilities as compared to normally developing individuals (Benson et al., 1993; Yirmiya, Solomonica-Levi, & Shulman, 1996; Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996; Zelazo et al., 1996), we were also interested in comparing the ToM abilities of individuals with MR to those of normally developing children. A nonsignificant difference would suggest that individuals with MR do not differ from normally developing children in their ToM abilities. Furthermore, a nonsignificant difference between individuals with MR and normally developing individuals, and significant differences between individuals with autism and both normally developing individuals and individuals with MR, would suggest that individuals with autism suffer from a unique impairment in performance on ToM tasks. However, if the ToM abilities of individuals with MR are significantly worse than those of normally developing individuals, then the reported deficit in ToM in individuals with autism cannot be described as a unique deficit in autism, as suggested by Baron-Cohen and others (Baron-Cohen, 1989b; Happé, 1995) but rather as characterizing other nonautistic individuals with MR as well.

The second aim of our study was to extend previous work by identifying the conditions under which the differences in performance on ToM tasks of individuals with autism, individuals with MR, and normally developing individuals may vary. Therefore, we examined 11 potential moderator variables. The first potential moderator variable was level of functioning (low vs. high) within the group or groups of individuals with autism. To examine the importance of the ToM deficit in autism, we

began by ascertaining whether this deficit characterizes individuals with autism regardless of their level of functioning.

The second potential moderator variable was the etiology of the individuals comprising the groups with MR. This potential moderator variable is important to the question regarding the uniqueness of the ToM deficit in autism as compared to MR, because of its role in defining specific comparison groups. In this field of study, different comparison groups such as individuals with known syndromes versus those with unknown (nonspecific) etiology were used to investigate the ToM hypothesis in autism. Thus, the specific characteristics of any given comparison group may relate to the results and the interpretation of any single study (Prior, Dahlstrom & Squires, 1990; Sigman, 1994; Yirmiya & Shulman, 1996; Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996). For example, individuals with Down syndrome are known to have high empathic abilities and better social skills compared to other individuals with MR of an unknown etiology (Beeghly, Weiss-Perry, & Cicchetti, 1990; Kasari, Mundy, Yirmiya, & Sigman, 1990). Thus, differences between individuals with autism and individuals with Down syndrome in ToM abilities may indicate a specific strength of the individuals with Down syndrome and not necessarily a deficiency in these skills in individuals with autism. In this case, a comparison group of individuals with MR of unknown etiology is crucial for testing the uniqueness of the ToM deficit in individuals with autism. To date, only one study (Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996) included two comparison groups, one with Down syndrome and the other with MR of unknown etiology, and the results varied depending on the specific group of individuals with MR to whom the group with autism was compared.

As mentioned above, ToM abilities of individuals with autism have been studied using different tasks. Therefore, the third potential moderator variable tested in the current meta-analyses was type of task. There were two reasons for examining type of task as a potential moderator variable. First, normally developing children pass some ToM tasks at age 2 ½ years, whereas other tasks are passed only after the age of 4 years. The wide variety of contents and contexts examined in the different tasks raises the possibility that these various tasks may not be testing the exact same abilities and that some are easier than others. The second reason is a methodological one: Because of the difficulty associated with recruiting and studying rare syndromes such as autism, many researchers who identify well-matched groups of participants administer several tasks to the same groups of individuals. Therefore, intergroup analyses within different types of tasks are important because they allow a more in-depth and independent examination of the data.

Thus, the first three potential moderator variables examined in the current meta-analyses were level of functioning of the group with autism, etiology of MR, and type of task. Additional potential moderator variables included indices of CA and mental abilities as well as other methodological variables. The potential moderator variables examined in the study were selected according to suggestions made by different authors in the field, frequency of their appearance in empirical literature, and theoretical and methodological considerations. These potential moderator variables and the theoretical and empirical considerations that led to their inclusion are described below.

To claim that a deficit in ToM is unique to autism as compared with other groups, the compared groups must be as similar as possible in every aspect other than those aspects unique to the syndrome. Therefore, the fourth potential moderator variable was the method of matching between the groups. Variables that may affect performance on any task, such as CA and MA, should be as closely matched as possible. Matching for CA and MA can be conducted at three different levels. Matching is least stringent when no assessment is carried out as when researchers assume that CA is indicative of MA for normally developing children without further testing of MA. At the next level, matching groups is carried out by matching the means and standard deviations of the two groups on a number of relevant variables. Finally, the most stringent approach to matching groups involves individually matching participants from the two groups on a one-to-one basis, thus controlling more tightly for the possible effect of variance.

A related issue is that the MA of participants in the various studies was assessed using different mental ability tests. In many studies, researchers used a limited verbal or performance test to determine the MA of the participants. Boucher (1989) noted that one commonly used verbal test, the British Picture Vocabulary Scale (BPVS), tests only recognition of single words. Whereas most individuals with autism have serious verbal communication difficulties, they do develop verbal knowledge of single words. It is plausible that single word recognition is correlated differently with other aspects of verbal and general ability for various clinical groups. Happé (1994c, 1995) reported a significant association between performance on the Sally and Anne task and Smarties task, on the one hand, and the comprehension subtest of the Wechsler Intelligence Scale for Children—Revised (Wechsler, 1974), on the other, whereas no such correlation emerged with other subtests, such as Digit Span or Block Design. Therefore, we examined the possibility that type of test used for evaluating MA moderates the linkage between group membership and ToM abilities.

Researchers testing ToM abilities in individuals with autism explored associations between ToM and CA and MA. Some researchers reported significant correlations between performance on tasks assessing ToM abilities and CA (Baron-Cohen, 1991, 1992; Leslie & Frith, 1988), others reported significant correlations between performance on these tasks and VMA (Charman & Baron-Cohen, 1992; Eisenmajer & Prior, 1991; Happé, 1994c, 1995; Leekam & Perner, 1991), and yet others (Prior et al., 1990) reported a significant correlation between performance on these tasks and both VMA and CA. Conversely, some researchers fail to find any association between performance on ToM tasks and CA (Baron-Cohen et al., 1985; Perner, Frith, Leslie, & Leekam, 1989). Because of the question of universality of the ToM deficit in autism, it is important to determine whether the deficit is exhibited by all or almost all individuals with autism regardless of their CA and mental abilities. Therefore, in the current study, in addition to coding the status of low versus high functioning, we coded CA, VMA, and performance mental age (PMA) for the group with autism as well as for the other groups as potential moderator variables.

In summary, our first aim in the current study was to compare the ToM abilities of individuals with autism, MR, and normal development, as reported in the literature, by using the procedure

of meta-analysis. Three meta-analyses were undertaken: one comparing the ToM abilities of individuals with autism to those of individuals with MR, one comparing the ToM abilities of individuals with autism to those of normally developing children, and one comparing the ToM abilities of individuals with MR to those of normally developing children. Our second goal was to explore within each meta-analysis the contribution of the abovementioned potential moderator variables that may account for variations among findings of previous studies.

Method

Sample of Studies

Studies were identified by using the PsychINFO, PsychLIT, ERIC, Social SciSearch, Dissertation Abstracts International, and Medline computerized systems up to and including 1997. The key words used for the search were *autism and theory of mind, autism and social-cognition, autism and social perspective-taking,* and *autism and emotional perspective-taking.* Relevant studies were also gathered from the reference lists of published journal articles. In addition, letters requesting relevant studies were sent to senior researchers in the field. The criteria for including a study were (a) publication in a journal or as a dissertation and (b) the inclusion of at least two of the three following groups: Individuals with autism, MR, and normal development. In studies that included a comparison group of normal 3-year-olds in addition to a group of normal 4-year-olds, only data regarding the performance of the 4-year-olds were entered. A sample of 40 reports that met the above criteria was obtained.

Moderator Variables

The 11 potential moderator variables with their mutually exclusive categories are described below. The categories of each potential moderator variable are summarized in Table 1. (The first variable indicated in Table 1 is only an indicator for the clinical group being compared: autism or MR.)

Level of Autism: Low-Functioning Versus High-Functioning

The groups of participants with autism were differentiated into two groups: higher functioning and lower functioning as determined by the mean MA/IQ and standard deviation of the group and the author's descriptions of the samples. High-functioning groups were those with a mean IQ of 65 or higher. Low-functioning groups were those with a mean IQ lower than 65. Examination of mean IQ and standard deviation of the groups revealed that most groups were characterized by an IQ lower than 65, but the standard deviation suggested the possibility that some higher functioning individuals were included. Because of lack of information, it was impossible to differentiate between groups including low-functioning individuals exclusively and those including both low- and high-functioning individuals. Furthermore, an IQ of 65 was chosen as a cutoff score to differentiate low- from high-functioning groups rather than the traditional score of 70 because there were several studies in which the mean performance IQ of the group was above 70 (indicating high functioning), yet the full scale score was between 65 and 70.

Etiology of MR

One or more of the following groups of individuals with MR were included in studies included in the meta-analyses: individuals with Down syndrome; individuals with MR of unknown etiology (i.e., the researchers specifically wrote that these individuals suffer from MR or mental handicap[1] not resulting from any identifiable syndrome or etiology);

Table 1

Potential Moderator Variables and Their Categories

Moderator	Variable
A. Clinical group	0. Autism
	1. MR—Down syndrome
	2. MR—unknown etiology
	3. MR—unaccounted etiology
	4. Other
B. Level of functioning of autism group	1. High
	2. Low
C. Comparison group	1. MR—Down syndrome
	2. MR—unknown etiology
	3. MR—undifferentiated etiology
	4. Other
	5. Normal
D. Type of task	1. False belief—level 1
	2. Smarties
	3. Deception
	4. False belief—level 2
	5. Other
	6. Ignorance
	7. Picture story
	8. Desire
	9. Average *d* all tasks in study
E. Type of matching	1. No matching
	2. Group matching
	3. Individual matching
	4. CA only
F. MA tests	1. Verbal
	2. Performance
	3. Verbal and Performance
	4. Other test or task
	5. No information or uncodeable
G. CA of experimental group	1. Up to 11, including
	2. 12–16
	3. 17+
H. VMA of experimental group	1. Up to 5, including
	2. 6 and above
	3. No information
I. PMA of experimental group	1. Up to 5, including
	2. 6 and above
	3. No information
J. CA of comparison group	1. 4–5
	2. 6–11
	3. 12–16
	4. 17+
K. VMA of comparison group	1. Up to 5, including
	2. 6 and above
	3. No information
L. PMA of comparison group	1. Up to 5, including
	2. 6 and above
	3. No information

Note. The order of variables and their classes appears in Tables 2, 3, and 4 under the heading Categorical Variables in the order presented here. MR = mental retardation; CA = chronological age; MA = mental age; VMA = verbal mental age; PMA = performance mental age.

---

[1] The terms *mental retardation* and *mental handicap* are used interchangeably. Most publications originating in the United States refer to individuals who meet the criteria for MR as individuals with MR, whereas most publications originating in Great Britain refer to these individuals as individuals with mental handicaps.

and individuals with MR with undifferentiated etiology (i.e., the researchers did not collect information regarding etiological factors,[2] and the only requirement for inclusion in the group with MR was for the individuals to meet criteria for MR or mental handicap). Thus, for example, in this last group, individuals with known syndromes such as Down syndrome and individuals with MR of unknown etiology were most likely grouped together. In addition, participants with specific language impairment, rather than with MR, were included in two studies (Leslie & Frith, 1988; Perner et al., 1989), whereas in one study the comparison group was composed of participants with learning disability and normally developing children (Prior et al., 1990). Therefore, in the final analyses, we included these three studies in the "other" category resulting in four levels for this variable: Down syndrome, MR of unknown etiology, MR of undifferentiated etiology, and other.

Type of Task

ToM tasks were classified as (a) false belief, level 1; (b) Smarties paradigm; (c) deception; (d) false belief, level 2; (e) other tasks that were administered in some studies but not frequently enough to be included in a separate category (including brain functions, appearance vs. reality, explanation of action cognition, and speech cognition terms); (f) ignorance tasks; (g) picture story tasks; and (h) desire tasks. Whenever a study included more than one task administered to the same samples, the effect sizes representing the various tasks were averaged for the purpose of intergroup analyses, which did not involve the moderation of type of task. The assigned code for the averaged effect size was nine.

Type of Matching

Matching between groups varied depending on whether matching was performed and by which method. The three categories entered for this moderator variable were no matching, group matching on the basis of means and standard deviations, and individual matching.

MA Tests

Studies were initially coded in terms of whether the researchers used a test of partial verbal, partial performance, comprehensive verbal, or comprehensive performance abilities or a combination of the abovementioned assessments. This coding resulted in few studies in each level. Therefore, in the final coding, studies were categorized on the basis of whether the researchers matched participants by using verbal tests; performance tests; verbal and performance tests; and other tests or tasks. (Some studies were categorized as uncodeable.)

Mean CAs and MAs

To examine whether differences in ToM abilities between groups are moderated by CA and MA, we categorized studies included in the meta-analyses according to participants' mean CA, VMA, and PMA. CA for the clinical groups included the following subclasses: 6–11 years, 12–16 years, 17 years and older, and no information. Because of the younger age of the normally developing children, an additional level was added for the category of CA of the comparison group, and the following subclasses were coded: 4–5 years, 6–11 years, 12–16 years, 17 years and older, and no information. Indices of mental age (VMA and PMA) involved the same subclasses for all participants and were coded as representing a MA of less than 5 years, or 6 years and above. (Some were coded as providing no information.)

Reliability

Each effect size included in the meta-analyses was coded on all 11 potential moderator variables by two independent coders. Every effect

size was assigned a value (refer to Tables 2, 3, and 4 for descriptions of these values) for each of the moderator variables, with values indicating the appropriate within moderator category in the order presented in the Method section. Coders coded three to six articles at a time, then met for discussion and assessment of reliability. Kappa coefficients calculated per category of moderator variable ranged from .74 to 1.00, with a mean of .92. All disagreements were discussed and resolved by the authors.

Multiple Effect Sizes From Single Samples

In the field of autism, the same sample of participants is frequently administered more than one task, thereby producing more than one effect size per comparison. Because the assumption of independence is critical for meta analysis, the issue of multiple effect sizes from single samples has to be acknowledged and addressed. To address this issue, under the same label we entered any two studies or results explicitly mentioned by the authors as having the same participant sample or any two publications that had at least 75% overlap in participants in at least one of the groups (as could be inferred from the participants' characteristics and in which the same task was administered), and we combined the results and entered them as one.[3] There were three cases in which two publications were coded as coming from a single study (Baron-Cohen, 1991, 1992; Baron-Cohen et al., 1985, 1986; Leslie & Frith, 1988; Perner et al., 1989), and two cases in which three publications were coded as a single study (Tager-Flusberg & Sullivan, 1994a, 1994b, 1995; Yirmiya & Shulman, 1996; Yirmiya, Solomonica-Levi, & Shulman, 1996; Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996), whereas the remaining studies were coded as independent.

In the current meta-analyses, 22 studies included one ToM task and thus yielded only one effect size for a specific group comparison, 10 studies yielded two effect sizes, 6 studies yielded three effect sizes, and 2 studies yielded four effect sizes. For the overall group comparisons all effect sizes from the same sample were averaged so that each study provided only one effect size. The effect sizes from the same sample were disaggregated whenever we reported on categorical analyses for type of task and etiology of MR.

Planned Analyses

Analyses were conducted using the D-STAT program (Johnson, 1989, 1993). Meta-analysis involves four steps. (For a more detailed overview of these steps, see Erel & Burman, 1995.) In the first step, the effect sizes derived from the individual studies are combined to determine a composite mean weighted effect size, d, and its significance level. This is achieved by coding the direction of the sign of each effect size. The direction of the sign indicates whether the individual outcome in the study was in the expected direction. Specifically, in the current study, we hypothesized that participants with MR and normally developing children would perform better than the participants with autism. Therefore, results that reflected a better performance of individuals with MR or normally developing children in comparison to individuals with autism, regardless of significance level, were assigned a positive sign. Results that did not suggest this direction were assigned a negative sign. Simi-

larly, results that reflected a better performance of the individuals with normal development in comparison to individuals with MR, regardless of significance level, were assigned a positive sign, and results that did not suggest this direction were assigned a negative sign. After being coded for direction, each result was converted into an effect size (d). Effect sizes represent the magnitude of an effect and may be derived from means and standard deviations, proportions, frequencies, and significance levels (p values). The standardized effect sizes from the independent studies were then combined into a composite mean weighted effect size whose significance was determined (i.e., the effect size is significant if the confidence interval does not include a zero). Because greater weight is given to studies with larger samples, this procedure corrects for biases associated with small sample sizes.

The second step in a meta-analytic procedure is to determine the homogeneity statistic Q. This homogeneity statistic indicates whether the composite mean weighted effect size found in the first step is consistent across the studies included in the meta-analysis. If so, this effect size can be considered as representative of the population from which it was drawn and free from the effect of moderator variables. Meta-analysis is considered complete when homogeneity is achieved.

However, because of the effects of moderator variables, the composite mean weighted effect size calculated from a large number of independent studies is usually not homogeneous. In this case, the third step in a meta-analysis, categorical model testing, is used to identify moderator variables, that is, variables that explain the inconsistency. To identify potential moderator variables, we divided studies into groups according to the different categories of a certain potential moderator variable. Categorical model testing is analogous to analysis of variance. It yields two relevant statistics: a within-class effect indicating whether the effect sizes within each moderator variable are homogeneous (Q_{wi}) and a between-classes effect, analogous to a main effect in analysis of variance, indicating whether significant differences exist between the categories of any given potential moderator variable (Q_b). This model-fitting procedure is repeated for as many moderator variables as the experimenter chooses to examine, until homogeneity is achieved.

The fourth step in meta-analytic research involves computation of contrasts between two homogeneous groups of studies within the same potential moderator variable. Differences in mean weighted effect sizes from two homogeneous categories imply the existence of a moderator variable (Hedges & Olkin, 1983). Computation of contrasts may be carried out following a significant between-class effect even when one or both classes being contrasted are not homogeneous. In this case, a significant contrast suggests that the analyses are somewhat exploratory and that it is difficult to know exactly what other factors may have contributed to the significant between-class effect.

In the current meta-analyses, we conducted categorical model testing procedures for the 11 abovementioned potential moderator variables within each of the three meta-analyses. These analyses are referred to as first-order categorical model testing procedures. When nonhomogeneous categories were revealed, we conducted second-order categorical model testing procedures to examine whether potential moderator variables explain the lack of homogeneity found within a certain class of a specific potential moderator variable. For example, does the etiology of the individuals with MR explain the nonhomogeneous results found for the general comparison between individuals with autism and individuals with MR (of all etiologies) on the classical false-belief task? Similar to the abovementioned example, second-order categorical model testing procedures were carried out especially within the nonhomogeneous levels of the potential moderator variable of type of task. Post hoc contrasts between different levels within any given category were interpreted only when each of the two contrasted levels included at least three effect sizes.

Results

Overall Group Comparisons

Individuals With Autism Versus Individuals With MR

The comparison between individuals with autism and individuals with MR included 31 reports representing 24 independent studies, thus yielding 24 summary effect sizes, which were used in analyses for all categories with the exception of the categories pertaining to etiology associated with MR (in which the relevant summary effect sizes were disaggregated into effect sizes representing all groups of MR studied by researchers) and type of task (in which the relevant summary effect sizes were disaggregated to represent all tasks investigated). Effect sizes, corresponding 95% confidence intervals, homogeneity statistics, and listing of categorical variables for each study included in this comparison are presented in Table 2. Of the 24 independent effect sizes, 17 effect sizes suggest that individuals with autism perform less well than individuals with MR, and 7 effect sizes suggest no difference. None of the effect sizes suggest that individuals with autism perform better on ToM tasks than individuals with MR. The effect sizes ranged from −.19 to 2.28. The composite weighted mean effect size of .84 with its 95% confidence interval of .70–.98 implies that individuals with autism perform significantly less well on tasks assessing ToM compared to individuals with MR.

Individuals With Autism Versus Normally Developing Individuals

The comparison between individuals with autism and normally developing individuals consisted of 26 reports, which represented 22 independent studies yielding 22 effect sizes. These effect sizes were used in all analyses with the exception of categorical analyses pertaining to the category of type of task in which the relevant effect sizes were disaggregated to fully represent the various tasks used by researchers. Effect sizes, corresponding 95% confidence intervals, homogeneity statistics, and listing of categorical variables for each study are presented in Table 3. As can be seen in that table, 15 effect sizes suggest that individuals with autism perform less well than individuals with normal development, and 7 effect sizes suggest no difference. None of the effect sizes suggest that individuals with autism perform better than individuals with normal development on tasks assessing ToM abilities. The effect sizes ranged from −.40 to 4.06. The composite weighted mean effect size of .88 with its 95% confidence interval of .74–1.03 implies that individuals with autism perform significantly less well than normally developing individuals on tasks assessing ToM abilities.

Individuals With MR Versus Normally Developing Individuals

The comparison between individuals with MR and normally developing individuals consisted of 19 reports representing 17 independent studies, which yielded 17 effect sizes. These 17 effect sizes were used in all categorical analyses with the exception of categorical analyses regarding the potential moderator variables of etiology associated with MR and type of task, for

Table 2

Summary of Effect Sizes, Homogeneity, and Study Characteristics: Individuals With Autism Versus Individuals With Mental Retardation

Study	Effect size ($d_{general}$)	95% confidence limits for d		r	p	Deviation	Homogeneity	Categorical variables (as listed in Table 1)
		Lower	Upper					
Baron-Cohen (1989a)	1.17	0.43	1.91	.51	.00	0.34	−0.85	0/2/3/5/2/3/2/2/2/3/2/2
Baron-Cohen (1989b)	1.66	0.64	2.68	.65	.00	1.01	−3.77	0/2/1/4/2/3/2/2/2/3/2/2
Averaged Baron-Cohen (1991/1992)	1.07	0.30	1.83	.48	.01	0.24	−0.35	0/2/3/9/2/3/2/2/2/3/1/1
Baron-Cohen (1991)	0.85	0.10	1.60	.40	.03	0.20	−0.27	0/2/3/2/2/3/2/2/2/3/1/1
Baron-Cohen (1991)	0.79	0.04	1.53	.37	.04	0.13	−0.12	0/2/3/1/2/3/2/2/2/3/1/1
Baron-Cohen (1992)	0.83	0.09	1.58	.39	.03	0.18	−0.22	0/2/3/3/2/3/2/2/2/3/1/1
Baron-Cohen & Goodhart (1994)	0.89	0.05	1.73	.42	.04	0.24	−0.30	0/2/3/6/2/1/2/1/3/3/1/3
Averaged Baron-Cohen et al. (1985/1986)	1.29	0.56	2.01	.55	.00	0.46	−1.49	0/2/1/9/1/3/2/1/2/2/1/1
Baron-Cohen et al. (1985)	1.69	0.89	2.48	.65	.00	1.05	−6.59	0/2/1/1/1/3/1/1/2/2/1/1
Baron-Cohen et al. (1986)	0.52	−0.15	1.20	.26	.12	−0.14	−0.16	0/2/1/7/1/3/2/1/2/2/1/1
Baron-Cohen et al. (1994)	2.28	1.36	3.20	.76	.00	1.64	−12.13	0/2/3/5/2/1/2/1/3/2/1/3
Charman & Baron-Cohen (1992)	0.89	0.15	1.63	.42	.02	0.24	−0.39	0/2/2/1/3/2/1/2/3/1/1
Charman & Baron-Cohen (1995)	0.55	0.10	1.01	.27	.02	−0.11	−0.21	0/2/2/2/3/2/1/2/3/1/1
U. Frith, Happé, & Siddons (1994)	0.44	−0.28	1.16	.22	.20	−0.22	−0.36	0/2/2/1/1/1/2/2/3/2/1/3
Happé (1994a)	1.25	0.47	2.02	.54	.00	0.60	−2.24	0/1/3/7/2/1/3/2/3/4/2/3
Hughes & Russell (1993)	1.37	0.81	1.93	.57	.00	0.73	−6.34	0/2/3/3/2/1/2/1/3/2/1/3
Averaged Oswald & Ollendick (1989)	1.31	0.35	2.28	.57	.01	0.48	−0.94	0/2/2/9/2/5/2/3/2/3/3/2
Oswald & Ollendick (1989)	−0.04	−0.92	0.84	−.02	.93	−0.71	−2.46	0/2/3/7/2/5/2/3/2/3/3/2
Oswald & Ollendick (1989)	2.57	1.39	3.75	.80	.00	1.93	−10.09	0/2/3/3/2/5/2/3/2/3/3/2
Oswald & Ollendick (1989)	0.57	−0.32	1.47	.29	.21	−0.08	−0.03	0/2/3/1/2/5/2/3/2/3/3/2
Averaged Ozonoff et al. (1991)	1.19	0.54	1.84	.52	.00	0.37	−1.16	0/1/4/9/3/2/2/2/3/2/2
Ozonoff et al. (1991)	0.71	0.10	1.33	.34	.02	0.06	−0.03	0/1/4/1/3/2/2/2/3/2/2
Ozonoff et al. (1991)	1.32	0.66	1.98	.56	.00	0.68	−3.92	0/1/4/4/3/2/2/2/3/2/2
Averaged Leslie & Frith (1988), Perner et al. (1989)	2.21	1.34	3.08	.75	.00	1.41	−9.79	0/2/4/9/2/1/2/2/3/2/2/3
Leslie & Firth (1988)	2.01	1.12	2.90	.72	.00	1.37	−9.02	0/2/4/1/2/1/2/2/3/2/2/3
Perner et al. (1989)	2.08	1.23	2.94	.73	.00	1.45	−10.93	0/2/4/2/2/1/2/2/3/2/2/3
Averaged Prior et al. (1990)	0.54	−0.09	1.18	.27	.09	−0.31	−0.90	0/2/4/9/3/3/1/2/2/2/2/2
Prior et al. (1990)	0.66	0.03	1.30	.32	.04	0.01	0.00	0/2/4/1/3/3/1/2/2/2/2/2
Prior et al. (1990)	0.00	−0.62	0.62	.00	1.00	−0.67	−4.43	0/2/4/6/3/3/1/2/2/2/2/2
Prior et al. (1990)	0.44	−0.19	1.07	.22	.17	−0.22	−0.48	0/2/4/2/3/3/1/2/2/2/2/2
Averaged Reed (1994)	0.65	−0.11	1.42	.32	.07	−0.20	−0.24	0/2/3/9/3/5/2/2/3/4/1/3
Reed (1994)	0.55	−0.21	1.31	.27	.13	−0.11	−0.08	0/2/3/6/3/5/2/2/3/4/1/3
Reed (1994)	−0.16	−0.91	0.58	−.08	.64	−0.84	−4.71	0/2/3/8/3/5/2/2/3/4/1/3
Reed (1994)	0.97	0.19	1.76	.45	.01	0.32	−0.63	0/2/3/1/3/5/2/2/3/4/1/3
Averaged Reed & Peterson (1990)	1.10	0.28	1.92	.49	.01	0.27	−0.39	0/2/3/9/2/3/2/2/2/2/3/3
Reed & Peterson (1990)	0.82	0.02	1.62	.39	.04	0.16	−0.16	0/2/3/6/2/3/2/2/2/2/3/3
Reed & Peterson (1990)	1.04	0.22	1.86	.47	.01	0.39	−0.87	0/2/3/1/2/3/2/2/2/2/3/3
Averaged Russell et al. (1991)	2.09	1.11	3.07	.73	.00	1.28	−6.38	0/2/1/9/2/2/2/3/2/3/3/3
Russell et al. (1991)	2.39	1.36	3.43	.78	.00	1.75	−11.00	0/2/1/1/2/2/2/3/2/3/3/3
Russell et al. (1991)	1.46	0.57	2.35	.60	.00	0.81	−3.19	0/2/1/3/2/2/2/3/2/3/3/3
Averaged Sodian & Frith (1992)	0.31	−0.28	0.91	.16	.29	−0.56	−3.24	0/2/2/9/2/1/2/2/3/3/1/3
Sodian & Frith (1992)	0.03	−0.56	0.61	.01	.93	−0.65	−4.58	0/2/2/3/2/1/2/2/3/3/1/3
Sodian & Frith (1992)	0.44	−0.29	1.16	.22	.24	−0.23	−0.36	0/2/2/1/2/1/2/2/3/3/1/3
Averaged Tager-Flusberg (1992)	0.28	−0.86	1.41	.15	.63	−0.57	−0.97	0/1/1/9/2/2/1/3/1/1/3/1
Tager-Flusberg (1992)	0.79	−0.39	1.96	.39	.18	0.13	−0.05	0/1/1/5/2/2/1/3/1/1/3/1
Tager-Flusberg (1992)	−0.38	−1.52	0.76	−.20	.51	−1.04	−3.18	0/1/1/8/2/2/1/3/1/1/3/1
Averaged Tager-Flusberg & Sullivan (1994a, 1994b, 1995)	0.5	0.01	1.08	.27	.04	−0.32	−1.25	0/1/2/9/2/3/3/2/3/3/2/3
Tager-Flusberg & Sullivan (1994a)	0.00	−0.52	0.52	.00	1.00	−0.68	−6.26	0/1/2/6/2/3/2/2/3/3/2/3
Tager-Flusberg & Sullivan (1994a)	0.59	0.05	1.13	.29	.03	−0.07	−0.06	0/1/2/8/2/3/2/2/2/3/2/2
Tager-Flusberg & Sullivan (1994a)	0.19	−0.34	0.71	.09	.48	−0.49	−3.18	0/1/2/1/2/3/2/2/2/3/2/2
Tager-Flusberg & Sullivan (1994a)	0.70	0.16	1.24	.34	.01	0.05	−0.03	0/1/2/5/2/3/2/2/2/3/2/2
Tager-Flusberg & Sullivan (1994b)	0.00	−0.80	0.80	.00	1.00	−0.67	−2.63	0/1/2/6/2/3/3/2/3/3/2/3
Tager-Flusberg & Sullivan (1994b)	0.17	−0.63	0.97	.09	.68	−0.50	−1.46	0/1/2/4/2/3/3/2/3/3/2/3
Tager-Flusberg & Sullivan (1995)	0.36	−0.18	0.89	.18	.19	−0.31	−1.24	0/1/2/1/2/3/2/2/3/3/2/3
Tan & Harris (1991)	−0.19	−0.78	0.40	−.10	.52	−0.87	−8.13	0/2/3/8/2/1/2/2/3/3/2/3
Grand Averaged Yirmiya (1996)	0.84	0.22	1.47	.39	.01	0.00	0.00	0/2/3/9/2/3/3/2/2/4/2/2
Averaged Yirmiya, MR-unknown (1996)	0.81	0.09	1.53	.38	.03	−0.08	−0.05	0/2/2/9/3/3/3/2/2/4/2/2
Yirmiya & Shulman (1996)*	0.42	−0.28	1.12	.21	.24	−0.25	−0.47	0/2/2/1/3/3/3/2/2/4/2/2
Yirmiya & Shulman (1996)*	0.61	−0.10	1.32	.30	.09	−0.05	−0.02	0/2/2/6/3/3/3/2/2/4/2/2
Yirmiya & Shulman (1996)*	0.67	−0.04	1.38	.32	.07	0.01	0.00	0/2/2/8/3/3/3/2/2/4/2/2

Table 2 (continued)

Study	Effect size ($d_{general}$)	95% confidence limits for d		r	p	Deviation	Homogeneity	Categorical variables (as listed in Table 1)
		Lower	Upper					
Yirmiya, Solomonica-Levi, & Shulman (1996)	0.48	−0.26	1.22	.24	.20	−0.18	−0.23	0/2/2/3/3/3/3/2/2/4/2/2
Averaged Yirmiya MR-Down (1996)	0.66	−0.07	1.40	.32	.08	−0.23	−0.37	0/2/2/9/2/3/3/2/2/4/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	0.54	−0.19	1.27	.27	.14	−0.12	−0.10	0/2/1/1/2/3/3/2/2/4/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	−0.24	−0.96	0.48	−.12	.51	−0.92	−6.15	0/2/1/6/2/3/3/2/2/4/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	1.32	0.53	2.11	.56	.00	0.67	−2.75	0/2/1/8/2/3/3/2/2/4/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	0.06	−0.66	0.79	.03	.86	−0.60	−2.58	0/2/1/2/3/3/3/2/2/4/2/2
Zubris (1994)	0.38	−0.04	0.81	.19	.08	−0.51	−4.92	0/2/2/1/2/3/2/1/1/3/1/1

* Data for this task were taken from Yirmiya and Shulman (1996) and Yirmiya, Solomonica-Levi, Shulman, and Pilowsky (1996).

which relevant effect sizes were disaggregated to represent fully the different groups of individuals with MR and type of tasks included in the literature. Effect sizes, corresponding 95% confidence intervals, homogeneity statistics, and listing of categorical variables for each study are presented in Table 4. As can be seen in that table, 5 of the 17 effect sizes suggest that individuals with MR perform less well than normally developing individuals, and 12 effect sizes suggest no difference. None of the effect sizes suggest that individuals with MR perform better than normally developing individuals on tasks assessing ToM abilities. The effect sizes representing comparisons between individuals with MR and normally developing individuals on ToM tasks ranged from −0.26 to +1.83. The composite weighted mean effect size of .45, with its 95% confidence interval of .29–.61, indicates that individuals with MR perform significantly less well than normally developing individuals on tasks assessing ToM abilities.

Overall Group Comparisons: Summary

The analyses regarding overall group differences reveal the following significant findings: (a) Individuals with autism perform significantly less well than individuals with MR on tasks assessing ToM abilities, (b) individuals with autism perform significantly less well than normally developing individuals on tasks assessing ToM abilities, and (c) individuals with MR perform significantly less well than normally developing individuals on tasks assessing ToM abilities. Cohen's (1977) criteria are that an effect size (in the metric of d) of .20 is small, .50 is medium, and .80 is large. Thus, because (a) significant differences emerged between individuals with MR and normally developing children, and because (b) the significant effect size yielded for the comparison between individuals with autism and normally developing children, $d = .88$, is larger than the significant effect size yielded for the comparison between individuals with MR and normally developing children, $d = .45$, it may be concluded (a) that individuals with autism and individuals with MR show a limited understanding of ToM as compared

to normally developing children and (b) that the impairment in understanding ToM is more severe in autism. This suggests, therefore, that the severity of the impairment in ToM abilities rather than the impairment itself is unique to autism.

Categorical Model Testing

Although significant findings emerged for the three above-mentioned comparisons, homogeneity was rejected in all three sets of data: individuals with autism versus individuals with MR, $Q_{w(23)} = 62.15, p < .0001$; individuals with autism versus normally developing individuals, $Q_{w(21)} = 103.74, p < .0001$; and individuals with MR versus normally developing individuals, $Q_{w(16)} = 41.91, p < .001$. Homogeneity was rejected in each data set even after the largest outlier was removed from each analysis. Therefore, to explore whether potential moderator variables can account for the heterogeneity, we carried out categorical model fitting procedures for the three data sets.

Individuals With Autism Versus Individuals With MR

Results of categorical model testing conducted for the comparison between the performance of individuals with autism and that of individuals with MR are presented in Table 5. The first potential moderator variable analyzed was level of functioning of the group with autism: low versus high functioning. As can be seen in Table 5, both low- and high-functioning individuals with autism perform significantly less well than individuals with MR, although homogeneity was revealed only for studies with high-functioning individuals. The between-class effect was nonsignificant, indicating that level of functioning of the group with autism is not a moderator variable in comparisons between the ToM abilities of individuals with autism and individuals with MR.

The next potential moderator variable examined was the etiology of the individuals with MR who were included in the comparison groups. Twenty-five rather than 24 effect sizes were included in this analysis because Yirmiya and colleagues (Yir-

Table 3
Summary of Effect Sizes, Homogeneity, and Study Characteristics: Individuals With Autism Versus Normally Developing Individuals

Study	Effect ($d_{general}$)	95% confidence limits for d		r	p	Deviation	Homogeneity	Categorical variables (as listed in Table 1)
		Lower	Upper					
Baron-Cohen (1989a)	1.43	0.70	2.17	.59	.00	0.57	−2.25	0/2/5/5/1/3/2/2/2/1/3/3
Baron-Cohen (1989b)	4.06	2.53	5.60	.90	.00	3.23	−16.90	0/2/5/4/1/3/2/2/2/2/3/3
Averaged Baron-Cohen (1991/1992)	1.04	0.32	1.75	.47	.00	0.16	−0.19	0/2/5/9/1/3/2/2/2/1/3/3
Baron-Cohen (1991)	0.94	0.23	1.64	.43	.01	0.09	−0.06	0/2/5/2/1/3/2/2/2/1/3/3
Baron-Cohen (1991)	0.90	0.20	1.60	.42	.01	0.05	−0.02	0/2/5/1/1/3/2/2/2/1/3/3
Baron-Cohen (1992)	0.54	−0.19	1.27	.27	.14	−0.32	−0.72	0/2/5/3/1/3/2/2/2/1/3/3
Averaged Baron-Cohen et al. (1985/1986)	1.87	1.19	2.55	.69	.00	1.03	−8.40	0/2/5/9/1/3/2/1/2/1/3/3
Baron-Cohen et al. (1985)	1.71	1.04	2.38	.66	.00	0.88	−6.39	0/2/5/1/1/3/1/1/2/1/3/3
Baron-Cohen et al. (1986)	1.69	1.02	2.35	.65	.00	0.86	−6.24	0/2/5/7/1/3/2/1/2/1/3/3
Bauminger (1996)	0.52	−0.10	1.15	.27	.10	−0.38	−1.36	0/1/5/4/2/3/1/2/2/2/2/2
Averaged Benson (1995)	0.39	−0.34	1.11	.19	.29	−0.52	−1.90	0/1/5/9/3/3/1/2/2/2/2/2
Benson (1995)	0.42	−0.31	1.14	.21	.26	−0.37	−0.96	0/1/5/1/3/3/1/2/2/2/2/2
Benson (1995)	0.17	−0.55	0.89	.09	.64	−0.62	−2.80	0/1/5/8/3/3/1/2/2/2/2/2
Averaged Bowler (1992)	0.00	−0.72	0.72	.00	1.0	−0.92	−6.10	0/1/5/9/1/3/3/2/2/4/3/3
Bowler (1992)	0.00	−0.72	0.72	.00	1.0	−0.88	−5.62	0/1/5/1/1/3/3/2/2/4/3/3
Bowler (1992)	0.00	−0.72	0.72	.00	1.0	−0.88	−5.62	0/1/5/4/1/3/3/2/2/4/3/3
Charman & Baron-Cohen (1992)	1.12	0.29	1.96	.50	.01	0.28	−0.41	0/2/5/2/1/3/2/1/2/1/3/3
U. Frith et al. (1994)	0.55	−0.11	1.20	.27	.09	−0.32	−0.88	0/2/5/1/1/1/2/2/3/1/1/3
Happé (1994a)	2.38	1.66	3.10	.77	.00	1.57	−17.23	0/1/5/7/1/1/3/2/3/2/3/3
Leekam & Perner (1991)	1.04	0.37	1.72	.47	.00	0.20	−0.31	0/2/5/1/1/3/2/2/2/1/3/3
Averaged Leekam & Prior (1994)	−0.40	−1.08	0.29	−.20	.26	−1.34	−13.83	0/2/5/9/1/1/4/2/3/5/3/3
Leekam & Prior (1994)	−0.30	−0.99	0.39	−.15	.39	−1.19	−11.19	0/2/5/3/1/1/4/2/3/5/3/3
Leekam & Prior (1994)	−0.30	−0.99	0.39	−.15	.39	−1.19	−11.19	0/2/5/1/1/1/4/2/3/5/3/3
Leekam & Prior (1994)	−0.17	−0.85	0.52	−.09	.63	−1.05	−8.84	0/2/5/4/1/1/4/2/3/5/3/3
Averaged Leslie & Thaiss (1992)	0.98	0.27	1.68	.45	.01	0.10	−0.07	0/2/5/9/1/1/2/2/3/1/1/3
Leslie & Thaiss (1992)	1.21	0.47	1.94	.53	.00	0.36	−0.90	0/2/5/2/1/1/2/2/3/1/1/3
Leslie & Thaiss (1992)	0.04	−0.64	0.71	.02	.91	−0.84	−5.78	0/2/5/2/1/1/2/2/3/1/1/3
Leslie & Thaiss (1992)	0.95	0.25	1.66	.44	.01	0.10	−0.08	0/2/5/1/1/1/2/2/3/1/1/3
Averaged Reed (1994)	1.35	0.70	2.01	.57	.00	0.49	−2.06	0/2/5/9/3/5/2/2/3/2/2/3
Reed (1994)	0.65	0.05	1.26	.32	.03	−0.21	−0.43	0/2/5/6/3/5/2/2/3/2/2/3
Reed (1994)	0.11	−0.48	0.70	.06	.71	−0.77	−6.32	0/2/5/8/3/5/2/2/3/2/2/3
Reed (1994)	2.45	1.67	3.23	.78	.00	1.63	−16.34	0/2/5/1/3/5/2/2/3/2/2/3
Averaged Reed & Peterson (1990)	1.77	1.00	2.54	.67	.00	0.92	−5.25	0/2/5/9/1/3/2/2/2/2/3/3
Reed & Peterson (1990)	1.52	0.65	2.39	.62	.00	0.68	−2.27	0/2/5/6/1/3/2/2/2/2/3/3
Reed & Peterson (1990)	1.86	0.94	2.78	.69	.00	1.02	−4.65	0/2/5/1/1/3/2/2/2/2/3/3
Roth & Leslie (1991)	0.99	0.26	1.72	.45	.00	0.14	−0.13	0/2/5/1/1/4/3/3/3/1/3/3
Averaged Russell et al. (1991)	1.91	0.99	2.83	.70	.00	1.05	−4.87	0/2/5/9/1/2/2/3/2/1/3/3
Russell et al. (1991)	2.50	1.49	3.52	.79	.00	1.67	−10.23	0/2/5/1/1/2/2/3/2/1/3/3
Russell et al. (1991)	0.96	0.15	1.77	.44	.02	0.11	−0.07	0/2/5/3/1/2/2/3/2/1/3/3
Averaged Sodian & Frith (1992)	0.68	0.10	1.26	.33	.02	−0.22	−0.52	0/2/5/9/1/1/2/2/3/1/3/3
Sodian & Frith (1992)	0.13	−0.43	0.70	.07	.64	−0.76	−6.58	0/2/5/3/1/1/2/2/3/1/3/3
Sodian & Frith (1992)	0.93	0.26	1.60	.43	.00	0.08	−0.05	0/2/5/1/1/1/2/2/3/1/3/3
Averaged Tager-Flusberg & Sullivan (1994a)	1.02	0.40	1.65	.46	.00	0.15	−0.20	0/1/5/9/1/3/2/2/2/2/3/3
Tager-Flusberg & Sullivan (1994a)	0.68	0.07	1.29	.33	.03	−0.18	−0.33	0/1/5/8/1/3/2/2/2/2/3/3
Tager-Flusberg & Sullivan (1994a)	1.03	0.40	1.66	.46	.00	0.18	−0.32	0/1/5/1/1/3/2/2/2/2/3/3
Tan & Harris (1991)	0.00	−0.59	0.59	.00	1.0	−0.89	−8.35	0/2/5/8/2/1/2/2/3/2/2/3
Averaged Yirmiya (1996)	1.74	0.92	2.55	.67	.00	0.88	−4.37	0/2/5/9/3/3/3/2/2/2/2/2
Yirmiya & Shulman (1996)*	2.87	1.88	3.86	.83	.00	2.04	−16.24	0/2/5/1/3/3/3/2/2/2/2/2
Yirmiya & Shulman (1996)*	1.09	0.35	1.84	.49	.00	0.25	−0.41	0/2/5/6/3/3/3/2/2/2/2/2
Yirmiya & Shulman (1996)*	0.43	−0.27	1.14	.22	.22	−0.43	−1.42	0/2/5/8/3/3/3/2/2/2/2/2
Yirmiya, Solomonica-Levi, & Shulman (1996)	1.30	0.51	2.09	.56	.00	0.46	−1.25	0/2/5/3/3/2/3/2/2/2/2/2
Zubris, 1994	0.14	−0.29	0.58	.07	.52	−0.84	−12.68	0/2/5/1/1/3/2/1/1/1/1/1

* Data for this task were taken from Yirmiya and Shulman (1996) and Yirmiya, Solomonica-Levi, Shulman, and Pilowsky (1996).

Table 4

Summary of Effect Sizes, Homogeneity, and Study Characteristics: Individuals With Mental Retardation Versus Normally Developing Individuals

Study	Effect size ($d_{general}$)	95% confidence limits for d		r	p	Deviation	Homogeneity	Categorical variables (as listed in Table 1)
		Lower	Upper					
Baron-Cohen (1989a)	0.26	−0.41	0.93	.13	.44	−0.20	−0.34	3/−/5/5/1/3/2/2/2/1/3/3
Baron-Cohen (1989b)	0.71	−0.20	1.61	.35	.12	0.20	−0.19	1/−/5/4/1/3/2/2/2/2/3/3
Averaged Baron-Cohen (1991/1992)	0.23	−0.44	0.90	.12	.50	−0.24	−0.46	3/−/5/9/1/3/2/1/1/1/3/3
Baron-Cohen (1991)	0.07	−0.60	0.74	.04	.84	−0.45	−1.71	3/−/5/2/1/3/2/1/1/1/3/3
Baron-Cohen (1991)	0.07	−0.60	0.74	.04	.84	−0.45	−1.71	3/−/5/1/1/3/2/1/1/1/3/3
Baron-Cohen (1992)	0.28	−0.44	1.00	.14	.45	−0.24	−0.41	3/−/5/3/1/3/2/1/1/1/3/3
Averaged Baron-Cohen et al. (1985/1986)	1.03	0.37	1.70	.47	.00	0.62	−3.09	1/−/5/9/1/3/1/1/1/1/3/3
Baron-Cohen et al. (1985)	0.01	−0.63	0.66	.01	.96	−0.51	−2.34	1/−/5/1/1/3/1/1/1/1/3/3
Baron-Cohen et al. (1986)	1.70	0.98	2.43	.66	.00	1.23	−10.62	1/−/5/7/1/3/1/1/1/1/3/3
Averaged Benson et al. (1993)	0.96	0.23	1.69	.44	.01	0.53	−1.94	2/−/5/9/2/3/3/3/2/2/3/2
Benson et al. (1993)	0.93	0.20	1.66	.43	.01	0.43	−1.31	2/−/5/1/2/3/3/3/2/2/3/2
Benson et al. (1993)	0.66	−0.05	1.37	.32	.07	0.16	−0.18	2/−/5/4/2/3/3/3/2/2/3/2
Charman & Baron-Cohen (1992)	0.20	−0.61	1.01	.10	.62	−0.32	−0.57	2/−/5/2/1/3/2/1/1/1/3/3
U. Frith et al. (1994)	0.11	−0.67	0.89	.06	.78	−0.41	−1.05	2/−/5/1/2/1/1/1/3/1/1/3
Happé (1994a)	1.83	1.05	2.61	.68	.00	1.44	−12.57	3/−/5/7/1/3/2/3/2/3/3
Averaged Reed (1994)	0.66	−0.11	1.42	.32	.07	0.21	−0.28	3/−/5/9/1/5/3/1/3/2/2/3
Reed (1994)	0.00	−0.75	0.75	.00	1.00	−0.52	−1.84	3/−/5/6/1/5/3/1/3/2/2/3
Reed (1994)	0.28	−0.47	1.03	.14	.43	−0.23	−0.36	3/−/5/8/1/5/3/1/3/2/2/3
Reed (1994)	1.08	0.29	1.87	.48	.00	0.58	−2.03	3/−/5/1/1/5/3/1/3/2/2/3
Averaged Reed & Peterson (1990)	0.66	−0.13	1.45	.32	.10	0.21	−0.27	3/−/5/9/1/5/1/3/3/2/3/3
Reed & Peterson (1990)	0.52	−0.26	1.30	.26	.19	0.01	0.00	3/−/5/6/1/5/1/3/3/2/3/3
Reed & Peterson (1990)	0.52	−0.26	1.30	.26	.19	0.01	0.00	3/−/5/1/1/5/1/3/3/2/3/3
Averaged Russell et al. (1991)	−0.26	−0.98	0.46	−.13	.48	−0.75	−3.92	1/−/5/9/1/5/2/3/3/1/3/3
Russell et al. (1991)	0.00	−0.72	0.72	.00	1.00	−0.53	−2.00	1/−/5/1/1/5/2/3/3/1/3/3
Russell et al. (1991)	−0.35	−1.07	0.38	−.18	.35	−0.88	−5.55	1/−/5/3/1/5/2/3/3/1/3/3
Averaged Sodian & Frith (1992)	0.38	−0.12	0.89	.19	.14	−0.08	−0.08	2/−/5/9/1/1/2/1/3/1/3/3
Sodian & Frith (1992)	0.13	−0.38	0.63	.06	.62	−0.41	−2.34	2/−/5/3/1/1/2/1/3/1/3/3
Sodian & Frith (1992)	0.45	−0.17	1.07	.22	.14	−0.06	−0.04	2/−/5/1/1/1/2/1/3/1/3/3
Averaged Tager-Flusberg & Sullivan (1994a)	0.54	−0.06	1.15	.27	.07	0.10	−0.10	2/−/5/9/1/3/2/2/2/2/3/3
Tager-Flusberg & Sullivan (1994a)	0.23	−0.36	0.83	.12	.43	−0.29	−0.88	2/−/5/8/1/3/2/2/2/2/3/3
Tager-Flusberg & Sullivan (1994a)	0.62	0.01	1.22	.30	.04	0.11	−0.12	2/−/5/5/1/3/2/2/2/2/3/3
Tan & Harris (1991)	0.19	−0.40	0.78	.10	.52	−0.33	−1.15	3/−/5/8/2/1/2/2/3/2/2/3
Grand Averaged Yirmiya (1996)	1.45	0.68	2.23	.60	.00	1.05	−6.64	3/−/5/9/2/3/3/2/2/2/2/2
Averaged Yirmiya MR-Unknown (1996)	1.32	0.56	2.09	.56	.00	0.92	−5.37	2/−/5/9/3/3/3/2/2/2/2/2
Yirmiya & Shulman (1996)*	1.78	0.96	2.60	.67	.00	1.30	−9.45	2/−/5/1/2/3/3/2/2/2/2/2
Yirmiya & Shulman (1996)*	0.16	−0.53	0.86	.08	.65	−0.36	−1.00	2/−/5/6/2/3/3/2/2/2/2/2
Yirmiya & Shulman (1996)*	1.40	0.63	2.17	.58	.00	0.91	−5.21	2/−/5/8/2/3/3/2/2/2/2/2
Yirmiya, Solomonica-Levi, & Shulman (1996)	0.71	−0.02	1.44	.34	.05	0.21	−0.30	2/−/5/3/3/3/3/2/2/2/2/2
Averaged Yirmiya MR-Down (1996)	1.42	0.62	2.22	.59	.00	1.02	−5.97	1/−/5/9/2/3/3/2/2/2/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	1.48	0.67	2.29	.61	.00	0.99	−5.67	1/−/5/1/2/3/3/2/2/2/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	1.26	0.48	2.04	.54	.00	0.77	−3.61	1/−/5/6/2/3/3/2/2/2/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	0.43	−0.30	1.15	.21	.25	−0.09	−0.05	1/−/5/8/2/3/3/2/2/2/2/2
Yirmiya, Solomonica-Levi, Shulman, & Pilowsky (1996)	1.29	0.50	2.07	.55	.00	0.80	−3.84	1/−/5/3/2/3/3/2/2/2/2/2
Zelazo et al. (1996)	0.93	0.09	1.78	.44	.03	0.43	−0.99	1/−/5/2/3/3/3/2/2/3/1/2
Zubris (1994)	−0.24	−0.66	0.19	−.12	.27	−0.81	−11.95	2/−/5/1/1/3/2/1/1/1/1/1

Note. MR = mental retardation.
* Data for this task were taken from Yirmiya and Shulman (1996) and Yirmiya, Solomonica-Levi, Shulman, and Pilowsky (1996).

miya, Solomonica-Levi, & Shulman, 1996; Yirmiya, Solo-monica-Levi, Shulman, & Pilowsky, 1996) included two comparison groups in their studies, one composed of individuals with Down syndrome and another group composed of individuals with MR of unknown etiology. Individuals with autism were found to perform significantly less well than all individuals with MR regardless of their etiologies. Homogeneity was achieved only for studies comparing individuals with autism to individuals with MR as a result of Down's syndrome and to individuals with MR of unknown etiology. The between-class effect was

Table 5

Tests of Categorical Models: Study Effect Sizes for Studies Comparing Individuals With Autism and Individuals With Mental Retardation

Variable and class	Between-class effect (Q_b)	n^a	Mean weighted effect size (d_{i+})	95% CI for d_{i+} Lower	95% CI for d_{i+} Upper	r^+	Homogeneity within class (Q_{wi})
Type of autism group	3.92						
High functioning		4 (4)	0.84	0.50	1.19	.39	4.26
Low functioning		20 (20)	0.84	0.69	0.99	.39	57.89****
Comparison group	15.39***						
MR—Down syndrome		5 (5)	1.17	0.78	1.56	.51	8.58
MR—unknown		7 (7)	0.52	0.31	0.73	.25	2.51
MR—undifferentiated		10 (10)	0.99	0.76	1.23	.44	26.62***
Other		3 (3)	1.15	0.75	1.55	.50	9.28*
Type of task	27.76****						
False belief, level 1		14 (14)	0.93	0.74	1.13	.42	94.29****
Smarties		5 (5)	0.79	0.51	1.08	.37	11.20*
Deception		6 (6)	0.88	0.58	1.17	.40	22.43***
False belief, level 2		3 (3)	1.01	0.56	1.47	.45	6.65
Other		4 (4)	1.09	0.72	1.46	.48	8.74
Ignorance		6 (6)	0.34	0.08	0.60	.17	6.14
Picture story		3 (3)	0.61	0.17	1.05	.29	4.75
Desire		5 (5)	0.28	−0.02	0.59	.14	11.77*
Type of matching	5.30						
No matching		3 (3)	0.87	0.45	1.29	.40	2.63
Group matching		18 (18)	0.84	0.68	1.00	.39	57.31****
Individual matching		3 (3)	0.80	0.41	1.19	.37	2.16
MA tests	2.11						
Verbal		8 (8)	0.88	0.64	1.13	.40	39.24****
Performance		2 (2)	1.32	0.58	2.06	.55	5.62
Verbal and performance		12 (12)	0.79	0.61	0.97	.37	14.08
No information		2 (2)	0.91	0.31	1.51	.41	1.11
CA (in years) of experimental group: Autism	1.92						
≤11		2 (2)	0.48	−0.07	1.03	.23	0.16
12–16		19 (19)	0.88	0.72	1.04	.40	57.92****
17+		3 (3)	0.80	0.44	1.16	.37	2.14
VMA (in years) of experimental group: Autism	3.00						
≤5		7 (7)	0.87	0.64	1.09	.40	20.13**
≥6		14 (14)	0.78	0.59	0.96	.36	33.40***
No information		3 (3)	1.32	0.73	1.90	.55	5.62
PMA (in years) of experimental group: Autism	7.49*						
≤5		2 (2)	0.37	−0.03	0.77	.18	0.03
≥6		12 (12)	0.99	0.79	1.20	.44	13.97
No information		10 (10)	0.81	0.60	1.03	.38	40.66****
CA (in years) of control group: MR	10.62*						
4–5		1 (1)	0.28	−0.86	1.41	.14	0.00
6–11		7 (7)	1.20	0.93	1.48	.52	19.39**
12–16		13 (13)	0.69	0.51	0.87	.33	30.94***
17+		3 (3)	0.90	0.49	1.31	.41	1.20
VMA (in years) of control group: MR	3.11						
≤5		11 (11)	0.78	0.59	0.97	.36	24.60*
≥6		9 (9)	0.84	0.61	1.07	.39	28.65****
No information		4 (4)	1.24	0.76	1.72	.53	5.79
PMA (in years) of control group: MR	3.60						
≤5		6 (6)	0.66	0.41	0.91	.31	6.57
≥6		6 (6)	1.02	0.72	1.31	.45	4.78
No information		12 (12)	0.88	0.68	1.09	.40	47.20****

Note. CI = confidence interval; MR = mental retardation; MA = mental age; CA = chronological age; VMA = verbal mental age; PMA = performance mental age.
a Number of effect sizes followed by number of studies from which these effect sizes were derived.
* $p < .05$. ** $p < .01$. *** $p < .005$. **** $p < .001$.

significant and contrast analyses between the two homogeneous classes revealed a significant difference, $\chi^2(3, N = 12) = 8.29$, $p < .05$, indicating that the mean weighted effect size associated with studies comparing individuals with autism to individuals with MR of unknown etiology is significantly smaller than the mean weighted effect size representing comparisons between individuals with autism and individuals with Down syndrome. Thus, the potential moderator variable of type of etiology associ-

ated with MR is a moderator variable in comparisons between the ToM abilities of individuals with autism and individuals with MR. Furthermore, it is interesting to note that the effect size representing the comparison between individuals with autism and individuals with MR of undifferentiated etiologies (i.e., most likely including individuals with and without Down syndrome in the same group) is significantly larger than the effect size representing comparisons between individuals with autism and individuals with MR of an unknown etiology, $\chi^2(3, N = 17) = 8.52, p < .05$, but not significantly different from the effect size representing comparisons between individuals with autism and individuals with Down syndrome $\chi^2(3, N = 15) = .61, p > .05$. These data support the hypothesis that individuals with autism perform significantly less well on ToM tasks compared to all other individuals with MR and most strikingly so when they are compared to individuals with Down syndrome and MR of undifferentiated etiologies. In summary, the etiology associated with MR is an important moderator variable in studies comparing the ToM abilities of individuals with autism and individuals with MR.

Next we examined the potential moderator variable of type of task. Forty-six effect sizes were included in this analysis representing the tasks examined by the different authors and included in this potential moderator variable. Categorical model testing of type of task revealed that individuals with autism performed significantly less well than individuals with MR on seven of the eight types of tasks (see Table 5). The between-class effect for type of task was significant, and homogeneity was reached for four of the eight tasks: false belief, level 2; ignorance tasks; picture story tasks; and other tasks. Computation of contrasts between the categories did not reveal any significant differences. These data suggest that individuals with autism perform less well than all other individuals with MR on most tasks and that the potential moderator variable of type of task is not a moderator variable and does not explain the heterogeneity found in studies comparing ToM abilities of individuals with autism to those of individuals with MR.

Of the remaining potential moderator variables, only CA of the individuals with MR was found to be a moderator variable. The between-class effect was significant, and contrast analyses revealed that the mean weighted effect size associated with studies in which the individuals with MR had a mean CA of 6–11 years was significantly larger than the mean weighted effect size associated with studies in which the individuals with MR had a mean CA of 12–16 years, $\chi^2(3, N = 20) = 9.59, p < .05$. One should note that both mean weighted effect sizes associated with the abovementioned contrast were not homogeneous; therefore, it is difficult to know exactly what other factors in addition to CA may have contributed to the significant between-class effect. None of the other potential moderator variables was identified as moderator variables (i.e., even when the between-class effect was significant, contrast analyses did not reveal any significant findings between classes involving at least three effect sizes).

Second-order categorical analyses: False belief, level 1. As can be seen in Table 5, the class of studies representing false belief, level 1 tasks was the only class within the category of type of task in which there was a sufficient number of studies to carry out meaningful second-order categorical model testing.

For tasks assessing false belief, level 1 all comparisons between individuals with autism and individuals with MR yielded significant results: The performance of individuals with autism was worse than that of individuals with Down syndrome, MR of unknown and undifferentiated etiologies, and other diagnoses. The between-class effect for the potential moderator variable of the etiology of the individuals with MR within tasks assessing false belief, level 1 was significant (see Table 6). Homogeneity was achieved for comparisons involving individuals with MR of unknown and undifferentiated etiologies and for other mixed groups (some of which included individuals with MR). Homogeneity was not achieved for the comparison between individuals with autism and individuals with Down syndrome, which yielded the largest effect size of all comparisons.

Computations of contrasts between the groups of studies revealed that the mean weighted effect size associated with false belief, level 1 studies in which individuals with autism were compared with individuals with MR of unknown etiologies was significantly smaller than the mean weighted effect size representing studies in which individuals with autism were compared with individuals with Down syndrome, $\chi^2(3, N = 8) = 11.88, p < .01$. The class of studies pertaining to comparisons involving individuals with Down syndrome was not homogeneous; therefore, it is difficult to know exactly which other factors may have contributed to the significant between-class effect. Yet, these results once again support our hypothesis that the etiology of the individuals with MR is an important moderator variable and that the most noticeable differences emerge when individuals with autism are compared to individuals with Down syndrome.

The mean VMA and PMA of the individuals with MR were also identified as moderator variables in comparisons between individuals with autism and individuals with MR on false belief, level 1 tasks. The mean weighted effect size associated with studies in which the mean VMA of the individuals with MR was up to 5 years was significantly smaller than the mean weighted effect size associated with studies in which the mean VMA of the individuals with MR was 6 years and above, $\chi^2(2, N = 11) = 7.38, p < .05$, or unknown, $\chi^2(2, N = 9) = 6.72, p < .05$. In addition, the mean weighted effect size associated with studies in which the mean VMA of the individuals with MR was 6 years and above was significantly larger than the mean weighted effect size associated with studies in which the mean VMA of the individuals with MR was unknown, $\chi^2(2, N = 8) = 7.88, p < .05$. Similarly, the mean weighted effect sizes associated with studies in which the mean PMA of the individuals with MR was up to 5 years and 6 years and above were significantly smaller than the mean weighted effect size associated with studies in which the mean PMA of the individuals with MR was not reported, $\chi^2(2, N = 10) = 8.45, p < .05$, and $\chi^2(2, N = 11) = 9.31, p < .01$, respectively. However, because some of the classes of studies were not homogeneous, these results should be treated as exploratory because it is unclear which other factors may have contributed to the significant between-class effects.

Summary. In summary, the etiology associated with MR was found to be a moderator variable in comparisons between individuals with autism and individuals with MR in the first- and second-order categorical analyses involving all ToM tasks and false belief, level 1 tasks. Comparisons between individuals

Table 6
Individuals With Autism Versus Individuals With Mental Retardation: Significant Second-Order Categorical Testing Within False Belief, Level 1 Tasks

Variable and class	Between-class effect (Q_b)	n^a	Mean weighted effect size (d_{i+})	95% CI for d_{i+}		r^+	Homogeneity within class (Q_{wi})
				Lower	Upper		
Etiology of MR	14.44***						
MR—Down syndrome		3 (3)	1.35	0.87	1.83	.56	9.36*
MR—unknown		5 (5)	0.40	0.14	0.66	.19	0.05
MR—undifferentiated		4 (4)	0.85	0.45	1.26	.39	0.70
Other		3 (3)	0.95	0.56	1.35	.43	6.82
VMA (in years) of MR	8.80*						
≤5		6 (6)	0.66	0.40	0.92	.32	9.50
≥6		5 (5)	1.25	0.92	1.58	.53	68.83****
No information		3 (3)	1.23	0.71	1.75	.52	7.16
PMA (in years) of MR	12.12***						
≤5		3 (3)	0.70	0.36	1.03	.33	8.11*
≥6		4 (4)	0.65	0.31	1.00	.31	0.08
No information		7 (7)	1.39	1.07	1.71	.57	73.99****

Note. MR = mental retardation; VMA = verbal mental retardation; PMA = performance mental age; CI = confidence interval.
[a] Number of effect sizes followed by number of studies from which these effect sizes were derived.
* $p < .05$. *** $p < .005$. **** $p < .001$.

with autism and individuals with MR of an unknown etiology yielded a significantly smaller effect size than comparisons between individuals with autism and individuals with MR related to Down syndrome. CA of the individuals with MR was found to be a moderator variable in the first-order categorical analyses: Comparisons involving studies in which the CA of the individuals with MR was between 6 to 11 years yielded a significantly larger effect size compared to comparisons involving studies in which the CA of the individuals with MR was between 12 to 16 years. VMA and PMA of the individuals with MR were found to be moderator variables in the second-order categorical analyses involving false belief, level 1 tasks. When the VMA and PMA of the individuals with MR were up to 5 years, the associated effect sizes were significantly smaller than when the VMA and PMA were unknown. The associated effect size representing studies in which the VMA of the individuals with MR was up to 5 years was significantly smaller than the effect sizes representing studies in which the VMA was 6 years and above or unknown. When the VMA was 6 years and above, the associated effect size was significantly larger than when the VMA was unknown, and finally, when the PMA was 6 years and above, the associated effect size was significantly smaller than when the PMA was unknown.

Individuals With Autism Versus Individuals With Normal Development

Categorical model testing results for comparisons involving individuals with autism and normally developing individuals are presented in Table 7. The categories of mean CA of the individuals with autism and mean VMA of the individuals with normal development were the only categories for which the between-class effect was significant and for which contrast analyses revealed a significant difference. For mean CA of the individuals with autism, the mean weighted effect size associated with studies in which the mean CA of the individuals with autism was

up to 11 years was significantly smaller than the mean weighted effect sizes associated with studies in which the mean CA of the individuals with autism were 12–16 years, $\chi^2(2, N = 18)$ = 11.95, $p < .005$, and above 17 years, $\chi^2(2, N = 7)$ = 14.82, $p < .001$. Thus, as individuals with autism are older, they most likely have lower cognitive abilities, and the difference between their performance and that of normally developing children becomes more noticeable.

For mean VMA of the individuals with normal development, the mean weighted effect size associated with studies in which the mean VMA of the normally developing children was up to 5 years was significantly smaller than the mean weighted effect size associated with studies in which the mean VMA of the normally developing individuals was unknown, $\chi^2(2, N = 17)$ = 14.12, $p < .0001$. Inspection of the studies in which the VMA of the normally developing children was unknown and not reported revealed that many researchers assumed that their MA was similar to their CA and therefore did not measure MA. In these studies, the participants with autism and those with normal development were matched on the basis of the MA of the individuals with autism and the CA of the normally developing children. It may be that the MA of these normally developing children was higher than their CA. Therefore, the difference between the performance of the individuals with autism and the normally developing children was larger. Alternatively, most of these normally developing children were 6 years or older; thus, most likely they mastered most tasks. Because some of the contrasted classes were not homogeneous, it is difficult to know what other factors may have contributed to the significant between-class effect. Nonetheless, the current analyses suggest that CA of the individuals with autism and VMA of the normally developing children are moderator variables and therefore should be taken into account and assessed in ToM studies comparing the abilities of individuals with autism and normally developing children.

Table 7

Tests of Categorical Models for Study Effect Sizes for All Studies Comparing Individuals With Autism and Normal Participants

Variable and class	Between-class effect (Q_b)	n^a	Mean weighted effect size (d_{i+})	95% CI for d_{i+} Lower	95% CI for d_{i+} Upper	r^+	Homogeneity within class (Q_{wi})
Type of autism group	6.62						
High functioning		5 (5)	0.85	0.55	1.15	.39	25.63****
Low functioning		17 (17)	0.89	0.73	1.06	.41	78.05****
Type of task	51.53****						
False belief, level 1		15 (15)	0.89	0.71	1.07	.41	82.28****
Smarties		3 (3)	0.93	0.50	1.35	.42	0.40
Deception		5 (5)	0.42	0.11	0.74	.21	11.83*
False belief, level 2		4 (4)	0.38	0.01	0.76	.19	25.89****
Other		2 (2)	1.20	0.72	1.68	.52	0.66
Ignorance		3 (3)	0.99	0.57	1.40	.44	2.66
Picture story		2 (2)	2.00	1.52	2.49	.71	1.94
Desire		5 (5)	0.27	−0.01	0.56	.13	3.10
Type of matching	10.26**						
No matching		17 (17)	0.94	0.77	1.11	.43	85.42****
Group matching		2 (2)	0.25	−0.18	0.68	.12	1.41
Individual matching		3 (3)	1.13	0.71	1.55	.49	6.65
MA tests	10.38*						
Verbal		6 (6)	0.63	0.37	0.90	.30	36.45****
Performance		1 (1)	1.91	0.99	2.83	.69	0.00
Verbal and performance		13 (13)	0.92	0.73	1.11	.42	56.91****
Other test or task		1 (1)	0.99	0.26	1.72	.44	0.00
No information		1 (1)	1.35	0.70	2.01	.56	0.00
CA (in years) of experimental group: Autism	16.29****						
≤11		3 (3)	0.19	−0.20	0.58	.09	4.13
12–16		15 (15)	0.94	0.77	1.12	.43	60.30****
17+		4 (4)	1.25	0.87	1.62	.53	23.01****
VMA (in years) of experimental group: Autism	3.43						
≤5		3 (3)	0.72	0.38	1.05	.34	18.63****
≥6		17 (17)	0.89	0.72	1.05	.40	79.33****
No information		2 (2)	1.34	0.77	1.92	.56	2.35
PMA (in years) of experimental group: Autism	18.69****						
≤5		1 (1)	0.14	−0.29	0.58	.07	0.00
≥6		13 (13)	1.15	0.94	1.36	.50	44.21****
No information		8 (8)	0.76	0.53	0.99	.36	40.84****
CA (in years) of control group: Normal	23.46****						
4–5		12 (12)	0.80	0.61	0.99	.37	40.98****
6–11		8 (8)	0.96	0.71	1.21	.43	39.30****
12–16		1 (1)	2.38	1.66	3.10	.77	0.00
17+		1 (1)	0.00	−0.72	0.72	.00	0.00
VMA (in years) of control group: Normal	15.98****						
≤5		3 (3)	0.41	0.09	0.73	.20	4.09
≥6		5 (5)	0.70	0.40	1.00	.33	16.46**
No information		14 (14)	1.13	0.94	1.33	.49	67.21****
PMA (in years) of control group: Normal	13.68***						
≤5		1 (1)	0.14	−0.29	0.58	.07	0.00
≥6		3 (3)	0.78	0.38	1.19	.37	7.11
No information		18 (18)	1.01	0.84	1.18	.45	82.94****

Note. CI = confidence interval; MA = mental age; CA = chronological age; VMA = verbal mental age; PMA = performance mental age.
[a] Number of effect sizes followed by number of studies from which these effect sizes were derived.
$*p < .05.$ $**p < .01.$ $***p < .005.$ $****p < .001.$

As can be seen in Table 7, individuals with autism were found to perform significantly less well than normally developing children on seven of the eight types of tasks: false belief, level one; Smarties; deception; ignorance; picture story; false belief, level 2; and other tasks. No significant differences between individuals with autism and normally developing individuals emerged for tasks assessing desire. Close examination of the articles pertaining to desire revealed that ceiling effects for the normally developing children were evident in only one study (Yirmiya,

Solomonica-Levi, Shulman, & Pilowsky, 1996) in which the normally developing individuals were rather old for this task (mean CA of 8 years; mean MA of 9 years, 2 months). Desire tasks are among the easiest ToM tasks and are passed by very young normally developing children (Astington & Gopnik, 1991; Wellman & Bartsch, 1994). All of the individuals with autism who participated in the desire tasks had a mean MA much higher than 3 years. Therefore, the lack of significant findings between normally developing children and individuals

with autism on desire tasks is most likely due to the task being relatively easy for both individuals with autism and normally developing children.

The between-class effect for the potential moderator variable of type of task was significant. However, contrast analysis revealed significant differences involving picture story tasks and desire; false belief, level 2; deception and false belief, level 1 tasks only. Picture story tasks, which appear in each of these contrasts, were administered in two studies only. Because our criteria for contrast analysis necessitated at least three independent studies, at this time, one cannot conclude that type of task is a moderator variable in comparison between individuals with autism and normally developing children. Perhaps as more studies are conducted, type of task will be found to be a moderator variable. These results suggest that individuals with autism perform significantly less well than normally developing children on most tasks assessing ToM abilities. On the basis of the current data, the potential moderator variable of type of task does not function as a moderator variable in comparisons between the ToM abilities of individuals with autism and normally developing children.

Homogeneity was achieved in four of the seven tasks for which significant differences were yielded: Smarties, ignorance, picture story, and other tasks. Homogeneity was achieved for desire tasks as well. Homogeneity was not achieved for the effect sizes assessing false belief, level 1 ToM tasks, for the effect sizes associated with deception tasks, and for the effect sizes assessing false belief, level 2. Therefore, second-order categorical model testing procedures of potential moderator variables were applied to the largest class, that of tasks assessing false belief, level 1 (which included 15 effect sizes).

Second-order categorical analyses: False belief, level 1. For false belief, level 1 tasks the second-order categorical model testing analyses revealed that CA and VMA of the normally developing children and type of matching are moderator variables. These findings are presented in Table 8. For CA of the normally developing children, the between-class effect was sig-

nificant. Contrast analyses revealed that the mean weighted effect size representing comparisons in which the normally developing children had a mean CA of 4 to 5 years was significantly smaller than the mean weighted effect sizes representing comparisons in which the normally developing children had a mean CA of 6 to 11 years, $\chi^2(2, N = 14) = 16.85, p < .0001$. For VMA of the normally developing children, the between-class effect was significant, and contrasts analyses revealed that the mean weighted effect size representing comparisons in which the normally developing children had a VMA of up to 5 years was significantly smaller than the mean weighted effect sizes representing comparisons in which the normally developing children had a mean VMA of 6 years and older, $\chi^2(2, N = 6) = 19.73, p < .0001$, or in which the VMA of the normally developing children was not reported, $\chi^2(2, N = 12) = 6.72, p < .05$. Regarding type of matching, the between-class effect was significant. Contrast analyses revealed that the mean weighted effect size representing comparisons in which the groups were not matched was significantly smaller than the mean weighted effect size representing comparisons in which the groups were matched on a one-to-one basis, $\chi^2(1, N = 15) = 13.44, p < .0005$. In all of the above contrast analyses within false belief, level 1 tasks, one or both classes of studies that were contrasted were not homogeneous. Therefore, these analyses are somewhat exploratory, and it is difficult to know exactly what other factors may have contributed to the significant between-class effects.

Summary. In summary, the results reveal that the performance of individuals with autism was inferior compared to that of normally developing children on most tasks assessing ToM abilities. CA of the individuals with autism and VMA of the normally developing children were found to be moderator variables in the first-order categorical analyses involving all ToM tasks. As the participants with autism were older, the difference between their performance and that of the normally developing children became more noticeable. Furthermore, the difference between individuals with autism and normally developing chil-

Table 8

Individuals With Autism Versus Individuals With Normal Development: Significant Second-Order Categorical Testing Within False Belief, Level 1 Tasks

Variable and class	Between-class effect (Q_b)	n^a	Mean weighted effect size (d_{i+})	95% CI for d_{i+}		r^+	Homogeneity within class (Q_{wi})
				Lower	Upper		
CA (in years) of control group: normal	23.13****						
4–5		10 (10)	0.75	0.55	0.96	.35	38.03****
6–11		4 (4)	1.73	1.31	2.14	.65	21.11****
12–16		1 (1)	0.00	−0.72	0.72	.00	0.00
VMA (in years) of control group: Normal	20.16****						
≤5		3 (3)	0.41	0.09	0.73	.20	3.90
≥6		3 (3)	1.69	1.23	2.16	.65	21.01****
No information		9 (9)	0.94	0.70	1.19	.43	37.20****
Type of matching	13.44***						
No matching		12 (12)	0.75	0.55	0.94	.35	47.83****
Individual matching		3 (3)	1.69	1.23	2.16	.65	21.01****

Note. CI = confidence interval; CA = chronological age; VMA = verbal mental age.
[a] Number of effect sizes followed by number of studies from which these effect sizes were derived.
*** $p < .005$. **** $p < .001$.

dren, although significant for all VMAs of the normally developing children, was significantly larger when the individuals with autism were compared to normally developing children with better verbal abilities. In the second-order categorical analyses involving false belief, level 1 tasks, CA and VMA of the normally developing children and type of matching were found to be moderator variables. Comparisons involving older normally developing children between the ages of 6–11 years yielded a significantly larger effect size compared to comparisons involving younger normally developing children. Similarly, when the VMA of the normally developing children was higher, a significantly larger effect size was yielded compared to when the VMA of the normally developing children was lower. Finally, a significantly larger effect size was yielded for comparisons involving studies in which participants with autism and normally developing children were group matched than when they were matched on a one-to-one basis.

Individuals With MR Versus Normally Developing Individuals

As can be seen in Table 9, all individuals with MR (those with Down syndrome and those with MR of unknown and undifferentiated etiologies) performed significantly worse than normally developing individuals. The between-class effect was significant, but contrast analyses did not reveal a significant difference between the mean weighted effect size associated with the different classes. Thus, etiology of MR was not found to be a moderator variable in comparisons involving individuals with MR and normally developing children.

Categorical modeling of the potential moderator variables of mean CA and PMA of the individuals with MR and mean CA and VMA of the normally developing children indicated that these are indeed moderator variables. Regarding CA of the individuals with MR, the mean weighted effect size representing studies in which the mean CA of the individuals with MR was between the ages of 12–16 years was significantly smaller than the mean weighted effect size representing studies in which the mean CA of the individuals with MR was 17 years and older, $\chi^2(2, N = 14) = 23.65, p < .0001$. Similarly, the mean weighted effect size representing studies in which the PMA of the individuals with MR was up to age 5 years was significantly smaller than it was in studies in which the PMA of the individuals with MR was 6 years and older, $\chi^2(2, N = 9) = 6.76, p < .05$. Thus, the difference between the performance of the individuals with MR and normally developing children is significantly larger when older rather than younger individuals with MR are included, most likely because of their lower mental abilities. In contrast, the findings regarding PMA suggest that more noticeable differences emerge between individuals with MR and normally developing children when the individuals with MR have a PMA of 6 years and higher than when their mean PMA is up to age 5. At first, these results may seem contradictory; however, inspection of these studies reveal that when the individuals with MR had a PMA of 6 years or higher, they were matched with older normally developing children compared to when they had a PMA of 5 years and lower and were matched with younger normally developing children. Thus, it appears that the CA and the VMA of the normally developing children

to whom the individuals with MR were matched explain the significant between-class effect yielded for PMA of the individuals with MR.

Indeed, regarding CA of the normally developing children, the mean weighted effect size representing studies in which the CA of the normally developing children was between the ages of 4 to 5 years was significantly smaller than the mean weighted effect size representing studies in which the mean CA of the normally developing children was 6 to 11 years, $\chi^2(2, N = 16) = 7.05, p < .05$. Regarding VMA of the normally developing children, the mean weighted effect size representing studies in which the VMA of the normally developing children was up to 5 years was significantly smaller than it was in studies in which the VMA of the normally developing children was not tested and reported, $\chi^2(2, N = 14) = 7.03, p < .05$. Thus, as normally developing children are older in age and therefore most likely have higher verbal abilities, the discrepancy between their performance on ToM tasks and the performance of individuals with MR on these tasks becomes more striking. When no information regarding the VMA of the normally developing children is provided because researchers assume that CA and MA are equivalent for normally developing children, the difference between normally developing children and individuals with MR is more noticeable. Categorical modeling regarding the remaining potential moderator variables did not reveal any additional moderator variables.

Although the between-class effect for type of task was significant, contrast analyses did not reveal any significant differences between classes including more than two effect sizes. Individuals with MR performed significantly less well than normally developing individuals on six of the eight types of tasks (see Table 9). Nonsignificant differences were yielded for the classes of studies in which the Smarties task and the deception tasks were used. Homogeneity was achieved for seven types of tasks (five of which were significant as well). These results suggest that individuals with MR perform significantly less well than normally developing children on most but not all tasks assessing ToM abilities and that the variable of type of task is not a moderator variable in comparisons between individuals with MR and normally developing children.

Second-order categorical analyses: False belief, level 1. Homogeneity was not achieved for false belief, level 1 tasks. Second-order categorical model testing of all other potential moderator variables was conducted for the class of studies using false belief, level 1 ToM tasks (see Table 10). As can be seen in Table 10, the potential moderator variables of CA of the individuals with MR, CA of the normally developing children, and type of matching were found to be moderator variables within false belief, level 1 tasks. No significant differences were revealed between individuals with MR and normally developing children when the individuals with MR were younger than 17 years. In contrast, a significant difference was found between individuals with MR who were 17 years old or older and normally developing children, with the individuals with MR performing significantly less well than the normally developing individuals. Contrast analyses yielded a significant difference between studies in which the individuals with MR were 17 years old and older and the remaining two subclasses, $\chi^2(2, N = 6) = 11.23, p < .005$, for the comparison with individuals with MR who

Table 9

Tests of Categorical Models for Study Effect Sizes for All Studies Comparing Individuals With Mental Retardation and Normally Developing Individuals

Variable and class	Between-class effect (Q_b)	n^a	Mean weighted effect size (d_{i+})	95% CI for d_{i+} Lower	95% CI for d_{i+} Upper	r^+	Homogeneity within class (Q_{wi})
Comparison group	6.41*						
MR—Down syndrome		5 (5)	0.75	0.40	1.09	.35	10.74
MR—unknown		7 (7)	0.24	0.01	0.47	.12	14.18*
MR—undifferentiated		6 (6)	0.54	0.25	0.82	.26	13.44*
Type of task	27.51****						
False belief, level 1		10 (10)	0.31	0.10	0.52	.15	28.22***
Smarties		3 (3)	0.34	−0.10	0.79	.17	2.65
Deception		4 (4)	0.25	−0.08	0.57	.12	8.29
False belief, level 2		2 (2)	0.68	0.12	1.24	.32	0.01
Other		2 (2)	0.46	0.01	0.90	.22	0.60
Ignorance		3 (3)	0.47	0.04	0.91	.23	2.73
Picture story		2 (2)	1.76	1.23	2.30	.66	0.05
Desire		4 (4)	0.40	0.07	0.72	.19	4.14
Type of matching	2.09						
No matching		12 (12)	0.38	0.20	0.57	.19	30.44***
Group matching		5 (5)	0.66	0.34	0.99	.31	9.37
MA tests	1.07						
Verbal		5 (5)	0.57	0.28	0.87	.28	14.21*
Verbal and performance		9 (9)	0.42	0.20	0.64	.21	22.73**
No information		3 (3)	0.32	−0.12	0.76	.16	3.90
CA (in years) of experimental group: MR	24.57****						
≤11		3 (3)	0.65	0.22	1.07	.31	3.14
12–16		9 (9)	0.16	−0.04	0.37	.08	8.56
17+		5 (5)	1.16	0.82	1.51	.50	5.65
VMA (in years) of experimental group: MR	5.40						
≤5		8 (8)	0.29	0.07	0.52	.14	14.28
≥6		6 (6)	0.72	0.44	1.01	.34	16.41*
No information		3 (3)	0.44	0.01	0.87	.21	5.82
PMA (in years) of experimental group: MR	6.95*						
≤5		4 (4)	0.16	−0.14	0.45	.08	10.02*
≥6		5 (5)	0.73	0.41	1.05	.34	5.96
No information		8 (8)	0.49	0.25	0.74	.24	18.99*
CA (in years) of control group: Normal	19.61****						
4–5		9 (9)	0.22	0.00	0.43	.11	15.14
6–11		7 (7)	0.68	0.41	0.95	.32	7.16
12–16		1 (1)	1.83	1.05	2.61	.68	0.00
VMA (in years) of control group: Normal	8.20*						
≤5		3 (3)	0.02	−0.32	0.36	.01	6.00
≥6		3 (3)	0.65	0.25	1.06	.31	6.38
No information		11 (11)	0.56	0.35	0.77	.27	21.32*
PMA (in years) of control group: Normal	17.77****						
≤5		1 (1)	−0.24	−0.66	0.19	−.12	0.00
≥6		2 (2)	1.19	0.66	1.72	.51	0.83
No information		14 (14)	0.50	0.31	0.68	.24	23.31

Note. CI = confidence interval; MR = mental retardation; MA = mental age; CA = chronological age; VMA = verbal mental age; PMA = performance mental age.
a Number of effect sizes followed by number of studies from which these effect sizes were derived.
* $p < .05$. ** $p < .01$. *** $p < .005$. **** $p < .001$.

are under the age of 12 years and $\chi^2(2, N = 7) = 20.84$, $p < .0001$, for comparisons involving individuals with MR who are between the ages of 12 and 16 years. As individuals with MR get older, the discrepancy between their CA and MA is bigger, their IQ is thus lower, their performance on ToM tasks is poorer, and the difference between their performance on ToM tasks and that of normally developing children becomes more noticeable.

For the potential moderator variable regarding the CA of the normally developing individuals, the between-class effect was significant and the subcategories were found to be homoge-

neous. Significant differences between individuals with MR and normally developing individuals were found for studies in which the normally developing individuals were older than 6 years in CA, but not when the normally developing children were under 6 years of age. Thus, when young normally developing children are matched with individuals with MR, there are no significant differences in their performance on ToM tasks, most likely as a result of matching on MA. The mean weighted effect size resulting from studies in which the normally developing children were younger than 6 years differed significantly from that asso-

Table 10

Individuals With Mental Retardation Versus Normally Developing Individuals: Significant Second-Order Categorical Testing Within False Belief, Level 1 Tasks

Variable and class	Between-class effect (Q_b)	n^a	Mean weighted effect size (d_{i+})	95% CI for d_{i+} Lower	95% CI for d_{i+} Upper	r^+	Homogeneity within class (Q_{wi})
CA (in years) of experimental group: MR	21.30****						
≤11		3 (3)	0.19	−0.23	0.61	.09	1.02
12–16		4 (4)	0.00	−0.29	0.28	.00	3.26
17+		3 (3)	1.24	0.79	1.69	.53	2.64
CA (in years) of control group: Normal	19.82****						
4–5		6 (6)	0.01	−0.24	0.26	.00	3.32
6–11		4 (4)	1.06	0.67	1.45	.47	5.05
Type of matching	8.95***						
No matching		7 (7)	0.14	−0.09	0.38	.07	10.66
Group matching		3 (3)	0.92	0.50	1.36	.42	8.60*

Note. CI = confidence interval; CA = chronological age; MR = mental retardation.
[a] Number of effect sizes followed by number of studies from which these effect sizes were derived.
*$p < .05$.   ***$p < .005$.   ****$p < .001$.

ciated with studies in which the normally developing children were between the ages of 6 to 11 years, $\chi^2(1, N = 10) = 19.82$, $p < .0001$. As normally developing children are older, their performance on ToM tasks improves. As a result, the difference between their performance on ToM tasks and that of individuals with MR is greater.

The potential moderator variable of type of matching was found to be a moderator variable. Contrast analysis revealed that the mean weighted effect size associated with studies in which no matching was conducted was significantly smaller than that associated with studies in which matching between participants was based on group matching, $\chi^2(1, N = 10) = 8.96$, $p < .005$. However, because one of the contrasted class of studies was not homogeneous, it is difficult to know what other factors (e.g., matching procedures on the basis of CA of the normally developing children with MA of the individuals with MR) may have contributed to the significant between-class effect for type of matching. Categorical modeling regarding the remaining potential moderator variables did not reveal any moderator variables involving classes of two or more effect sizes.

Summary. In summary, CA of both normally developing children and individuals with MR were identified as moderator variables in the first- and second-order analyses and need to be taken into account in future research. As individuals with MR get older, the difference between their performance on ToM tasks and that of normally developing children becomes more noticeable. Similarly, as normally developing children get older, their ToM abilities improve, and the difference between their performance on ToM tasks and that of individuals with MR becomes more noticeable. In addition, PMA of the individuals with MR and VMA of the normally developing children were identified as moderator variables in the first-order categorical analyses indicating that when individuals with MR are matched with normally developing children whose mental abilities are assumed to be higher than those of the individuals with MR, the difference in ToM abilities becomes more noticeable. Finally, type of matching was found to be a moderator variable in the

second-order categorical analyses, suggesting that matching procedures should be carefully considered in studying the ToM abilities of these individuals.

Summary of Categorical Model Testing

For the comparison between individuals with autism and those with MR, the first-order categorical analyses revealed two moderator variables: the etiology of the individuals with MR and mean CA of the individuals with MR. Second-order categorical analyses for false belief, level 1 tasks revealed once more that the etiology associated with MR is an important moderator variable in comparisons involving individuals with autism and individuals with MR. In addition, mean VMA and PMA of the individuals with MR were identified as moderator variables.

For the comparison between individuals with autism and normally developing children, first-order categorical analyses regarding all ToM tasks revealed that mean CA of the individuals with autism and VMA of the normally developing children are moderator variables. Second-order categorical analyses regarding false belief, level 1 tasks revealed that CA and VMA of the normally developing children and type of matching are moderator variables.

For the comparison between individuals with MR and normally developing children, first-order categorical testing for all ToM tasks revealed that CA and PMA of the individuals with MR and CA and VMA of the normally developing children are moderator variables. Second-order categorical analyses within false belief, level 1 tasks revealed that CA of the individuals with MR and of the normally developing children and type of matching are moderator variables.

Thus, there seems to be no justification to group together individuals with MR regardless of their etiology. Furthermore, CA and indices of MA and matching procedures are important moderator variables that need to be taken into account when comparing ToM abilities of individuals with autism, individuals with MR, and normally developing children.

Discussion

Summary of Meta-Analytic Review

The direction and the magnitude of the difference between ToM abilities of individuals with autism, nonautistic individuals with MR, and normally developing individuals were examined in a series of three meta-analyses. Our focus was on whether individuals with autism consistently perform less well than all other individuals and whether nonautistic individuals with MR perform less well on ToM tasks compared to normally developing individuals. In addition, we explored the conditions under which the magnitude and direction of the difference within each comparison varied by identifying potential moderator variables.

The difference between the ToM abilities of individuals with autism and those with MR was significant, $d = .84$, as were the differences between the ToM abilities of individuals with autism and normally developing individuals, $d = .88$, and between the ToM abilities of individuals with MR and normally developing children, $d = .45$. These data confirm that the deficit in ToM abilities characterizes individuals with autism but is not unique to autism because it is manifested by individuals with MR as well. What may be unique to autism is the severity of the impairment rather than the impairment itself. Different moderator variables were found for the three sets of data including etiology associated with MR, CA of all three groups, PMA and VMA of the individuals with MR, VMA of the normally developing children, and type of matching in comparisons involving normally developing children.

Although type of task was not found to be a moderator variable, significant differences between individuals with autism and other individuals, and between individuals with MR and normally developing children, emerged for some but not all types of tasks. It may be that various tasks measure different aspects of the same underlying construct, with some tasks being easier than others. Ungerer (1989) suggested that tasks conceptualized as sharing a common deficit are not always as highly correlated as one would expect. Thus, in addition to various difficulty levels, it may be that various ToM tasks require somewhat different combinations of abilities, with some tasks relying more on linguistic abilities and others relying more on cognitive and social–emotional abilities. In the meta-analyses, we focused on intergroup comparisons and thus cannot map questions of specific task difficulties for each group of participants. Future research with normally developing children and with clinical samples may provide a better mapping of the different tasks and abilities associated with passing or failing them. Therefore, researchers should factor out different potentially contributing elements, leaving only understanding mental states, and then examine differences among the groups. Similarly, level of functioning of the individuals with autism was not found to be a moderator variable. This finding does not mean that low- and high-functioning individuals with autism do not differ in their ToM abilities. In the context of comparisons to individuals with MR and normally developing children, however, level of functioning of the individuals with autism is not a moderator variable because of matching on CA and MA/IQ between the individuals with autism and the comparison group or groups (i.e., higher functioning individuals with autism were matched with higher functioning individuals without autism, whereas lower functioning individuals with autism were matched with lower functioning individuals with MR or with younger normally developing children).

Implications

The first implication is that ToM deficits can no longer be conceptualized as a core deficit that is unique to autism. Second, researchers interested in autism have typically included groups of individuals with MR as their comparison groups. When studying ToM abilities, the most common comparison group chosen by researchers has been some group of individuals with MR, many times without a detailed account of the specific etiologies of the individuals with MR. Our findings suggest that the selection of comparison groups must be carefully evaluated and taken into consideration when interpreting data and inferring conclusions. Many researchers in the field of autism use a comparison group of individuals with Down syndrome because this syndrome is easily identified and diagnosed, and it is one of the most common causes for MR. Yet persons with Down syndrome display a unique profile: They show specific strengths in attentional, social, and emotional abilities (Beeghly et al., 1990; Kasari et al., 1990; Mundy, Sigman, Kasari, & Yirmiya, 1988). When etiology of MR was considered, the effect size yielded for comparisons between individuals with Down syndrome and individuals with autism was larger than that yielded for comparisons of individuals with autism and individuals with MR of unknown or undifferentiated etiologies. Thus, the common practice of including individuals with Down syndrome as a comparison group referred to as individuals with MR, or the practice of including individuals with Down syndrome and individuals with MR resulting from other etiologies in one comparison group, may not be the best approach in research examining abilities and impairments of individuals with MR in general and in studying impairments in individuals with autism, in particular. In the first article on autism and ToM, Baron-Cohen et al. (1985) reported that 86% of their participants with MR, diagnosed with Down syndrome (who had a mean VMA of 2 years and 11 months) passed the Sally and Ann task, whereas 80% of the participants with autism (who had a higher mean MA) failed this task. This finding may be as important in its contribution to understanding the strengths displayed by individuals with Down syndrome as in its contribution to understanding the impairments associated with autism.

The ToM hypothesis in autism has evoked great enthusiasm and hope that the unique and underlying core deficit in autism finally has been discovered. One aim of the meta-analyses was to examine to what extent this deficit is exhibited by individuals with MR as well, thus enabling us not only to question the uniqueness of the deficit in autism but also to further enhance our understanding of the underlying components of this ability and its origins.

It is important to note that in addition to the current meta-analyses, researchers exploring ToM abilities in individuals with schizophrenia and in deaf children (Corcoran, Frith, & Mercer, 1995; C. D. Frith & Corcoran, 1996; Peterson & Siegal, 1995) suggested that other clinical groups may also be impaired in ToM ability. The findings that individuals with MR and schizo-

phrenia and deaf children are also impaired in their ToM abilities does not exclude the possibility that distinct elements (empathic ability, various dimensions of cognitive ability, social relations, etc.) of this ability are differently impaired in various groups of individuals. For example, the studies on deaf children point to the importance of social learning or of an acquired element in ToM abilities, whereas studies regarding individuals with MR point to the importance of cognitive faculties. By further examining the ToM abilities of different clinical groups, we may enhance our understanding of its varying components and their origin. For example, individuals with attention deficit/hyperactivity disorder (ADHD) who have been tested as a comparison group in examining "executive functions" abilities of individuals with autism have not yet been tested on tasks assessing ToM abilities. This group of children with ADHD may share some of the attentional difficulties experienced by individuals with autism and thus provide important information regarding the role of attention mechanisms in ToM abilities, which are currently missing in the literature.

Indices of CA were identified as important moderator variables in comparisons involving all groups. CA of the individuals with autism was found to be a moderator variable in comparisons between individuals with autism and normally developing children. The most striking difference was revealed when older, rather than younger, participants with autism were included in the studies. None of the MA indices of the individuals with autism was found to be a moderator variable. Thus, individuals with autism, regardless of their MA, have a deficit in their ToM abilities, and older participants show the most severe deficit compared to normally developing children.

CA and indices of MA of the individuals with MR were found to be moderator variables: CA, PMA, and VMA in comparisons between individuals with autism and individuals with MR and CA and PMA in comparisons between individuals with MR and normally developing children. As individuals with MR have higher VMA and PMA, the difference between their performance on false belief, level 1 tasks and that of the individuals with autism becomes more striking in favor of the individuals with MR. In comparisons with normally developing children, older participants with MR and those matched with normally developing children who were more competent showed the most severe deficits. The fact that indices of MA of the individuals with MR play a role in comparisons involving individuals with MR but that indices of MA of individuals with autism do not play a role in comparisons involving individuals with autism point once more to the severity of the deficit in ToM abilities associated with autism.

CA and VMA of the normally developing children were found to be moderator variables. Older normally developing children and those with higher VMA showed the greatest strength in their ToM abilities compared to the groups of individuals with autism and MR. Most researchers on the development of ToM in normally developing children suggest that by age 4 years, children are able to know about mental states of others. Therefore, one would not expect such diverse findings between comparisons involving normally developing children of different ages, all of whom are older than 4 years of age. Two hypotheses may be offered for this finding. First, it may be that the important factor is not the CA of the normally developing individuals per

se, but rather their CA in comparison to the CA and MA of the clinical group (with autism or MR) to whom they are compared. Many researchers assume that CA and MA correspond in normally developing children and therefore, they match normally developing individuals with individuals with MR or autism on the basis of the CA of the normally developing individuals and the MA of the individuals with MR or autism. These matching procedures may result in an underestimation of the VMA of the normally developing children, which in reality may be higher than their CA, thus explaining why CA of the normally developing children was found to be a moderator variable.

Second, CA and indices of MAs of normally developing individuals (even long after the age of 4 years) and of individuals with MR may be important factors in comparisons involving these individuals. Manifestations of abilities and impairments are likely to change with psychological and physical development (Sigman, 1994). The characteristic picture in normal development and in MR may include changes in ToM abilities, whereas in autism the deficit may be invariable and thus does not change dramatically enough to allow for significant differences to emerge as a function of MA. This possibility once more points to the severity of the ToM deficit in autism.

Happé (1995) found a relation between CA and ToM abilities for the individuals with MR, with older participants performing worse than younger ones, but attributed the effect to sampling error. In our study, CA of individuals with MR was identified as a moderator variable substantiating the results reported by Happé from the perspective of intergroup comparisons rather than an intragroup perspective.

Finally, type of matching was identified as a moderator variable in comparisons involving normally developing children. Most researchers assume that CA of normally developing children is identical to their MA and therefore do not test the children with developmental or IQ tests. The current results strongly suggest that matching procedures should be carefully considered when comparing the ToM abilities of normally developing children with those of individuals with autism or mental retardation.

Study Limitations

The first limitation of the current meta-analyses is the lack of a sufficient number of studies that would enable an examination of the effect of a larger number of potential moderator variables. For example, many studies were published on ToM abilities of individuals with autism, yet only five studies included individuals with Down syndrome. Because of the paucity of studies including individuals with Down syndrome, it was impossible to examine whether potential moderator variables account for the lack of homogeneity within these studies. In addition, the small number of studies did not allow a full examination of interaction or additive effects. For example, it may be that the developmental trajectories of individuals with different etiologies of various ages and mental abilities differ. Future studies should focus on different clinical groups and include individuals of various abilities within each group.

Gender is an important potential moderator variable that could not be examined in the current meta-analyses because researchers did not provide separate data for male and female individuals. Because ToM abilities are to some extent a social phenome-

non, it would have been interesting to examine the effects of gender on ToM abilities of individuals with normal development, MR, and autism. We hope that this variable is examined in future research. Other variables not examined in previous studies, such as age of diagnosis, severity of symptoms associated with autism, as well as intervention programs received by the participants, may also have important implications for studying ToM abilities.

Another factor possibly limiting the current analyses is that we relied on journal articles and dissertations. It may be that additional nonsignificant results were not included in published articles or dissertations. This issue, known as the "file drawer problem," is problematic in all research and with meta-analyses in particular. However, most researchers were interested in the comparison between individuals with autism on the one hand and individuals with MR and normal development on the other hand. Because the comparison between individuals with MR and normal development was not their main focus, they would have probably published their data regardless of whether or not significant differences emerged between these groups of individuals. Thus, the finding of the current meta-analyses regarding the significant difference between individuals with MR and normally developing individuals seems robust.

Finally, there were not enough studies to examine the contribution of CA, MA, and IQ independently. IQ is a standardized norm referenced (age corrected) score. Because many of the individuals with autism and MR who are older than 16 years cannot be tested with the adult versions of most intelligence tests, they are administered the younger versions, which are appropriate for their abilities but do not provide norms for their age. Thus, researchers are faced with the problem of using the test that is most suitable for the individuals being tested (e.g., Wechsler Intelligence Scale for Children—3rd edition; Wechsler, 1991; for an adult with MR) but with no norms for interpretation of the performance in terms of IQ. At best, the researcher is able to assign a MA level implying that a 32-year-old individual with autism is functioning at a level similar to that expected from a normally developing 9-year-old. Yet, there may be crucial differences in abilities between an individual who is 32 years old and who functions at an MA level of 9 years and an individual who is 20 years old and also functions at an MA level of 9 years. What is missing here is the IQ; however, there is no valid way to infer IQ for participants who, on the one hand, are older than the ages for which the test was designed, even though the specific test is appropriate for their mental abilities, and yet on the other hand do not have the minimal required abilities to be tested on the test appropriate for their CA. In the current meta-analyses, MAs were coded because most researchers reported MAs. Yet, if more studies were available, it would have been possible to code IQ in addition to CA and MAs.

Conclusions

In the last 13 years, the conceptualization of autism as involving a unique cognitive deficit in ToM abilities was the focus of theoretical consideration and empirical activity. The current series of meta-analyses revealed that although individuals with autism manifest a more severe impairment in their ToM abilities than do individuals with MR, this deficit is also manifested in

individuals with MR. This implies that the deficit is not unique to autism, although it may be that individuals with autism and individuals with MR fail ToM tasks for different reasons.

Part of the initial enthusiasm surrounding the deficit in ToM abilities in individuals with autism may have resulted from the connotations that the term ToM evokes. The mind is characterized by cognitive and affective attributes and may be conceptualized as representing soul and brain. According to Webster's Encyclopedic Unabridged Dictionary of the English Language (1989), the first three definitions for mind, n. are as follows:

1. (in a human or other conscious being) the element, part, substance, or process that reasons, thinks, feels, wills, perceives, judges, etc.: the processes of the mind. 2. Psychol. the totality of conscious and unconscious mental processes and activities of the organism. 3. intellect or understanding, as distinguished from the faculties of feeling and willing; intelligence. (p. 911)

The first and third definition are clearly conflicting: In the first definition, the adjectives feels and wills are included, and the definition seems to incorporate both soul and brain, whereas in the third definition, mind is referred to as that which is different from feeling and willing, namely referring to intelligence. Thus, this third definition is much more in accordance with the brain rather than with the soul. Theories regarding the core deficit or deficits in autism are typically either of a cognitive or an affective–social nature. It may be that ToM was so appealing to those interested in autism because it seemed to capture feeling and thinking, affect and cognition, soul and brain. Although the term mind involves affect as well, the research paradigms testing ToM abilities in individuals with autism actually tap only cognition rather than normally flowing behaviors. Thus, these paradigms are more congruent with the third rather than the first definition of mind. Feldman (1992) supported this idea: "The literature on ToM is about children's knowledge of the social world, particularly of the mental states of the people they encounter" (p. 107). Similarly, Astington and Olson (1995) stated that "most attention has been devoted to understanding knowledge, belief, and perception" (p. 182) and that "social interaction is undoubtedly the context in which children acquire their understanding of people's minds, and information regarding children's participation in such interaction is a vital source of data in any attempt to explain the development of this understanding" (p. 183).

The dichotomy between studying cognition versus affect in general, or studying the construction of concepts versus the internalization of social understanding, in particular, exists not only in studying autism but also in the literature regarding the development of normal children. Yet, some researchers interested in ToM in normally developing children have begun to call for the integration of the two perspectives acknowledging that given a normal brain and appropriate social experience, the construction of the mind is universal (Astington & Jenkins, 1995; Astington & Olson, 1995; Brown, Donelan-McCall, & Dunn, 1996; Bruner, 1990; Dunn, 1996; Lalonde & Chandler, 1995). Researchers are beginning to describe associations between ToM abilities and social and emotional development (Astington & Jenkins, 1995; Dunn, 1995; Jenkins & Astington, 1996). Researchers should continue to explore in more depth the mutual and separate developmental courses of perceptual,

cognitive, and emotional states and how they relate to ToM development. Yet the behavioral manifestations of autism relate to the mind, to the most "human" and the most complex of all human behaviors: communication and social interactions. An attempt to reduce these human characteristics to either a purely cognitive or a purely affective impairment may thus be impossible.

The findings of the current meta-analyses suggest that the impairment in ToM in autism and in other clinical groups should not be conceptualized as an "all or none" phenomena. Future studies may assist in clarifying which groups of individuals, at what CAs, and of which mental (e.g., cognitive, social, affective, linguistic) abilities, pass different tasks assessing various aspects (e.g., cognitive, affective, linguistic) of ToM. It is only if we adopt the first definition of *mind* into our investigations of autism that we may be in the right direction to explain the enigma.

References

References marked with an asterisk indicate studies included in the meta-analysis.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology, 9,* 7–31.

Astington, J. W., Harris, P. L., & Olson, D. (1988). *Developing theories of mind.* New York: Cambridge University Press.

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

Astington, J. W., & Olson, D. R. (1995). The cognitive revolution in children's understanding of mind. *Human Development, 38,* 179–189.

*Baron-Cohen, S. (1989a). Are autistic children "behaviourists"? An examination of their mental–physical and appearance–reality distinctions. *Journal of Autism and Developmental Disorders, 19,* 579–600.

*Baron-Cohen, S. (1989b). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, 30,* 285–297.

*Baron-Cohen, S. (1991). The development of a theory of mind in autism: Deviance and delay? *Psychiatric Clinics of North America, 14,* 33–51.

*Baron-Cohen, S. (1992). Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology and Psychiatry, 33,* 1141–1155.

Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *Current Psychology of Cognition, 13,* 513–552.

Baron-Cohen, S. (1995). The eye-direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41–59). Hillsdale, NJ: Erlbaum.

*Baron-Cohen, S., & Goodhart, F. (1994). The "seeing-leads-to-knowing" deficit in autism: The Pratt and Bryant probe. *British Journal of Developmental Psychology, 12,* 397–401.

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

*Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4,* 113–125.

Baron-Cohen, S., & Ring, H. (1994). A model of the mindreading system: Neuropsychological and neurobiological perspectives. In C. Lewis & P. Mitchell (Eds.), *Origins of an understanding of mind* (pp. 183–207). Hillsdale, NJ: Erlbaum.

*Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). Recognition of mental state terms: Clinical findings in children with autism and a functional neuroimaging study of normal adults. *British Journal of Psychiatry, 165,* 640–649.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. (1993). *Understanding other minds: Perspectives from autism.* Oxford, England: Oxford University Press.

*Bauminger, N. (1996). *Loneliness, friendship and theory of mind in high-functioning children with autism.* Unpublished doctoral dissertation, University of California, Los Angeles.

Beeghly, M., Weiss-Perry, B., & Cicchetti, D. (1990). Beyond sensorimotor functioning: Early communicative and play development of children with Down syndrome. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental approach* (pp. 329–368). New York: Cambridge University Press.

*Benson, G. L. (1995). *Theory of mind and pragmatic performance in high-functioning children with autism.* Unpublished doctoral dissertation, University of Wisconsin—Madison.

*Benson, G., Abbeduto, L., Short, K., Bibler-Nuccio, J., & Maas, F. (1993). Development of theory of mind in individuals with MR. *American Journal on Mental Retardation, 98,* 427–433.

Boucher, J. (1989). The theory of mind hypothesis of autism: Explanation, evidence and assessment. *British Journal of Disorders of Communication, 24,* 181–198.

*Bowler, D. M. (1992). "Theory of mind" in Aspergers syndrome. *Journal of Child Psychology and Psychiatry, 33,* 877–893.

Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings and mothers. *Child Development, 67,* 836–849.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

Burack, J. A. (1992). Debate and argument: Clarifying developmental issues in the study of autism. *Journal of Child Psychology and Psychiatry, 33,* 617–621.

Butterworth, G., Harris, P. L., Leslie, A. M., & Wellman, H. M. (1991). *Perspectives on the child's theory of mind.* Oxford, England: Oxford University Press.

Chandler, M. J., Fritz, A. S., & Hala, S. (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., & Baron-Cohen, S. (1992). Understanding drawings and beliefs: A further test of the metarepresentational theory of autism. *Journal of Child Psychology and Psychiatry, 33,* 1105–1112.

*Charman, T., & Baron-Cohen, S. (1995). Understanding photos, models, and beliefs: A test of the modularity thesis of theory of mind. *Cognitive Development, 10,* 287–298.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Corcoran, R., Frith, C. D., & Mercer, G. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research, 17,* 5–13.

Dunn, J. (1995). Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. *Cognition and Emotion, 9,* 187–201.

Dunn, J. (1996). Children's relationships: Bridging the divide between cognitive and social development. *Journal of Child Psychology and Psychiatry, 37,* 507–518.

Eisenmajer, R., & Prior, M. (1991). Cognitive linguistic correlates of "Theory of mind" ability in autistic children. *British Journal of Developmental Psychology, 9,* 351–364.

Erel, O., & Burman, B. (1995). Interrelatedness of marital relations

and parent–child relations: A meta-analytic review. *Psychological Bulletin, 118,* 108–132.

Feldman, C. F. (1992). The new theory of theory of mind. *Human Development, 35,* 107–117.

Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin & P. Pufall (Eds.), *Piaget's theory: Prospects and possibilities* (pp. 107–134). Hillsdale, NJ: Erlbaum.

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

Frith, C. D., & Corcoran, R. (1996). Exploring "theory of mind" in people with schizophrenia. *Psychological Medicine, 26,* 521–530.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford, England: Blackwell Publishers.

*Frith, U., Happé, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development, 3,* 108–124.

Frye, D., & Moore, C. (1991). *Children's theories of mind: Mental states and social understanding.* Hillsdale, NJ: Erlbaum.

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

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.

Gopnik, A., & Slaughter, V. (1991). Young children's understanding of changes in their mental states. *Child Development, 62,* 98–110.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledging theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development, 62,* 83–97.

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

Happé, F. G. E. (1994b). *Autism: An introduction to psychological theory.* London: UCL Press.

Happé, F. G. E. (1994c). Wechsler IQ profile and theory of mind in autism: A research note. *Journal of Child Psychology and Psychiatry, 35,* 1461–1471.

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. (1991). The work of imagination. In A. Whiten (Ed.), *Natural theories of mind* (pp. 283–304). Oxford, England: Basil Blackwell.

Hedges, L. V., & Olkin, I. (1983). Clustering estimates of effect magnitude from independent studies. *Psychological Bulletin, 93,* 563–573.

Hobson, P. (1993). Understanding persons: The role of affect. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds* (pp. 204–226). Oxford, England: Oxford University Press.

Hobson, R. P. (1990a). Concerning knowledge of mental states. *British Journal of Medical Psychology, 63,* 199–213.

Hobson, R. P. (1990b). On acquiring knowledge about people and the capacity to pretend: Response to Leslie (1987). *Psychological Review, 97,* 114–121.

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

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.

Johnson, B. L. (1989). *DSTAT: Software for the meta-analytic reviews of research literatures.* Hillsdale, NJ: Erlbaum.

Johnson, B. L. (1993). *DSTAT 1.10: Software for the meta-analytic review. Upgrade documentation.* Hillside, NJ: Erlbaum.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2,* 217–250.

Kasari, C., Mundy, P., Yirmiya, N., & Sigman, M. (1990). Affect and attention in children with Down syndrome. *American Journal on Mental Retardation, 95,* 55–67.

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social–emotional consequences of coming early or late to a first theory of mind. *Cognition and Emotion, 9,* 167–185.

*Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40,* 203–218.

*Leekam, S. R., & Prior, M. (1994). Can autistic children distinguish lies from jokes? A second look at second-order belief attribution. *Journal of Child Psychology and Psychiatry, 35,* 901–915.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review, 94,* 412–426.

Leslie, A. M. (1988). Some implications of pretense for mechanisms underlying the child's theory of mind. In J. W. Astington, O. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 19–46). Cambridge, England: Cambridge University Press.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). New York: Cambridge University Press.

*Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology, 6,* 315–324.

Leslie, A. M., & Roth, D. (1993). What autism teaches us about meta-representation. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds* (pp. 83–111). Oxford, England: Oxford University Press.

*Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43,* 225–251.

Lewis, C., & Mitchell, P. (1994). *Origins of an understanding of mind.* Hillsdale, NJ: Erlbaum.

Lewis, C., & Osbourne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficits or linguistic artifact? *Child Development, 61,* 93–109.

Lillard, A. S. (1993). Young children's conceptualization of pretend: Action or mental representational states? *Child Development, 64,* 372–386.

Lillard, A. S. (1996). Body or mind: Children's categorizing of pretense. *Child Development, 67,* 1717–1734.

Moore, C. (1996). Theories of mind in infancy. *British Journal of Developmental Psychology, 14,* 19–40.

Moses, L. J., & Chandler, M. J. (1992). Traveler's guide to children's theories of mind. *Psychological Inquiry, 3,* 286–301.

Mundy, P., Sigman, M., Kasari, C., & Yirmiya, N. (1988). Nonverbal communication skills in Down syndrome children. *Child Development, 59,* 235–249.

*Oswald, P. D., & Ollendick, T. H. (1989). Role taking and social competence in autism and mental retardation. *Journal of Autism and Developmental Disorders, 19,* 119–127.

*Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive functioning deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32,* 1081–1105.

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

*Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Explora-

tion of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60,* 689–700.

Peterson, C. C., & Siegal, M. (1995). Deafness, conversion and theory of mind. *Journal of Child Psychology and Psychiatry, 36,* 459–474.

*Prior, M., Dahlstrom, B., & Squires, T. (1990). Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry, 31,* 587–601.

*Reed, T. (1994). Performance of autistic and control participants on three cognitive perspective-taking tasks. *Journal of Autism and Developmental Disorders, 24,* 53–66.

*Reed, T., & Peterson, C. (1990). A comparative study of autistic participants' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders, 20,* 555–567.

*Roth, D., & Leslie, A. M. (1991). The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Developmental Psychology, 9,* 315–330.

Ruffman, T., Olson, T., Ash, T., & Keenan, T. (1993). The ABCs of deception: Do young children understand deception in the same way as adults? *Developmental Psychology, 29,* 74–87.

*Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows" task as a measure of strategic deception in preschoolers and autistic participants. *British Journal of Developmental Psychology, 9,* 331–349.

Sigman, M. (1994). What are the core deficits in autism? In S. H. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain functioning* (pp. 139–157). New Jersey & London: Erlbaum.

Sigman, M. (1996). Behavioral research in childhood autism. In M. Lenzenwager & J. Hougaard (Eds.), *Frontiers of developmental psychopathology* (pp. 190–206). New York: Oxford University Press.

Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology, 9,* 173–188.

*Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 33,* 591–605.

*Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Development, 63,* 161–172.

*Tager-Flusberg, H., & Sullivan, K. (1994a). Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry, 35,* 1059–1075.

*Tager-Flusberg, H., & Sullivan, K. (1994b). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders, 24,* 577–586.

*Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics, 16,* 241–256.

*Tan, J., & Harris, P. L. (1991). Autistic children understand seeing and wanting. *Development and Psychopathology, 3,* 163–174.

Taylor, M. (1996). Social cognitive development from a theory of mind perspective. In R. Gelman, T. Au, E. C. Carterette (Gen. Ed.), & M. P. Friedman (Gen. Ed.), *Handbook of perception and cognition: Vol.*

13. Perceptual and cognitive development (pp. 283–329). San Diego, CA: Academic Press.

Ungerer, J. A. (1989). The early development of autistic children: Implication for defining primary deficits. In G. Dawson (Ed.), *Autism: Nature, diagnosis, and treatment* (pp. 75–91). New York: Guilford Press.

Webster's encyclopedic unabridged dictionary of the English language. (1989). New York: Gramercy Books.

Wechsler, D. (1974). *WISC-R Manual: Wechsler Intelligence Scale for Children—Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—3rd ed. (WISC—III).* San Antonio, TX: Psychological Corporation.

Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

Wellman, H. M. (1993). Early understanding of mind: The normal case. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 10–39). Oxford, England: Oxford University Press.

Wellman, H. M., & Bartsch, K. (1994). Children's early understanding of mind: Origins and development. In C. Lewis & P. Mitchell (Eds.), *Before belief: Children's early psychological theory* (pp. 331–354). Hove, England: Erlbaum.

Whiten, A. (1991). *Natural theories of mind, evaluation, development and simulation of everyday mindreading.* Oxford, England: Basil Blackwell.

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.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders, 9,* 11–30.

*Yirmiya, N., & Shulman, C. (1996). Seriation, conservation, and theory of mind abilities in individuals with autism, mental retardation, and normal development. *Child Development, 67,* 2045–2059.

*Yirmiya, N., Solomonica-Levi, D., & Shulman, C. (1996). The ability to manipulate behavior and to understand manipulation of beliefs: A comparison of individuals with autism, mental retardation and normal development. *Developmental Psychology, 32,* 62–69.

*Yirmiya, N., Solomonica-Levi, D., Shulman, C., & Pilowsky, T. (1996). Theory of mind abilities in individuals with autism, Down syndrome, and mental retardation of unknown etiology: The role of age and intelligence. *Journal of Child Psychology and Psychiatry, 37,* 1003–1014.

*Zelazo, P. D., Burack, J. A., Benedetto, E., & Frye, D. (1996). Theory of mind and rule use in individuals with Down's syndrome: A test of the uniqueness and specificity claims. *Journal of Child Psychology and Psychiatry, 37,* 479–484.

*Zubris, D. K. (1994). *Comparison of knowledge of self and other in children with autism, children with other developmental disabilities and preschoolers without disabilities.* Unpublished doctoral dissertation, George Washington University, Washington, DC.