Articles

Is there a Gender Difference in False Belief Development?

Tony Charman, *Institute of Child Health* and Ted Ruffman and Wendy Clements, *University of Sussex*

Abstract

The contribution of children's social environment to their acquisition of theory of mind skills, combined with the well documented advantage for girls in mental state talk with siblings, peers and mothers, might lead to a female advantage on false belief tasks. We present a post-hoc analysis of large datasets from two independent laboratories. A slight advantage for girls on false belief task performance was found in both datasets and was only apparent in younger but not older children. Language ability could be controlled for only in a smaller subsample of one dataset and cannot be ruled out as a potential mediator of this effect. However, if there is an age-specific advantage for girls in false belief acquisition it is a weak effect only.

Keywords: theory of mind; false belief; gender

Research into individual differences in theory of mind performance has shown several effects of the child's social environment. Children with siblings show an advantage over singletons (Perner, Ruffman, & Leekam, 1994), this advantage may be restricted to the presence of older but not younger siblings (Ruffman et al., 1998), and may only benefit verbally less able children (Jenkins & Astington, 1996). In addition, contact with extended family members may confer an additional advantage over the presence of siblings (Lewis et al., 1996). Parenting style is also associated with theory of mind ability (Ruffman, Perner, & Parkin, 1999), and this effect shows some gender specificity (Hughes, Deater-Deckard, & Cutting, 1999). Measures of attachment security are also associated with theory of mind ability, although the direction of any causal association remains to be determined (Fonagy, Redfern, & Charman, 1997; Meins et al., 1998). It has been suggested that the facilitative effects of (older) siblings, family members, parenting style and attachment relationships may operate via opportunities for shared experiences of pretend play and deception, and talk about feelings and internal mental states. Indeed, some specific evidence exists that co-operative sibling interaction and frequency of mental state talk in dyadic play with friends is associated

Please address correspondence to either: Tony Charman, Behavioural and Brain Sciences Unit, Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, UK, E-mail: t.charman@ich.ucl.ac.uk, or: Ted Ruffman, Experimental Psychology, University of Sussex, Brighton, East Sussex, BN1 9QG, UK. E-mail: *tedr@biols.susx.ac.uk*

longitudinally with false belief performance (Dunn et al., 1991; Hughes & Dunn, 1998). Further, the effects of such social interactions are not merely artefacts of language ability, although this clearly plays a role (Dunn et al., 1991; Hughes & Dunn, 1998; Jenkins & Astington, 1996).

We know that gender differences exist in the amount and type of emotion and mental state talk that occurs between parents (particularly mothers) and children, and between child dyads. Mothers talk more to girls than to boys and more of this talk consists of supportive speech acts, although this advantage may be restricted to toddlerhood rather than the pre-school and early school age years (Leaper, Anderson, & Saunders, 1998). Mothers talk more to daughters than sons about emotions at age 2, and the child's own level of emotion talk positively influences this association (Cervantes & Callanan, 1998; Dunn, Bretherton, & Munn, 1987). Older siblings also mention feeling states more frequently to girls than boys, and both mother and sibling emotion and mental state talk is associated with increased feeling state talk in 2-year-old girls compared to boys (Brown, Donelan-McCall, & Dunn, 1996). The association between early sibling interaction and later emotion understanding is stronger for girls than for boys (Brown & Dunn, 1996).

Combining these two strands of research might lead us to expect an advantage for girls over boys on theory of mind performance, at least in the earlier stages of acquisition. Is there any evidence to support such a thesis? Some studies have shown an advantage for girls on measures of mental state talk (Hughes & Dunn, 1998) and emotion understanding (Dunn et al., 1991). Few studies have investigated gender influence on theory of mind development as their primary aim. In a rare exception, Banerjee (1997) found that girls outperformed boys in one of two conditions on an emotion 'appearance-reality' task and on a task measuring understanding of emotion display rules. Likewise, Bosacki and Astington (1999) found that 11-year-old girls were better at assessing story characters' motives and feelings than were boys. There is some evidence of a female advantage in mindreading abilities in adulthood (Baron-Cohen & Hammer, 1997; Baron-Cohen et al., 1997) but other studies have failed to replicate this female advantage (Jarrold et al., 2000). However, these studies did not examine false belief understanding, considered to be the 'litmus test' of theory of mind reasoning. Most studies that report a post-hoc gender comparison show no significant advantage for girls or boys (e.g. Holmes, Black, & Miller, 1996; Jenkins & Astington, 1996), with only a few exceptions (Cutting & Dunn, 1999; Happé, 1995).

However, if we assume that any theory of mind gender effect would be expected to be a weak effect only, it may be that most previous research has lacked the statistical power to demonstrate such an effect. A power analysis, employing a t-test calculation with a variance inflation factor to model a power calculation for a sequential logistic regression (Cohen, 1988; Hsieh, Bloch, & Larsen, 1998), demonstrates that a weak effect would only have a 51% chance of being detected with alpha set at .05 in a total sample of 200 boys and 200 girls with age entered first into the regression equation. Power only reaches 88% with samples of 1000 children (500 boys, 500 girls). Theory of mind research does not usually employ samples larger than 50 subjects (though see Hughes and Cutting (1999) for an exception where in 119 same-sex twin pairs no theory of mind advantage was found for girls). We wished to take advantage of the large pool of false belief data residing in our laboratories in order to examine whether post-hoc analysis of a sufficiently large sample would reveal a theory of mind advantage for girls that might be expected from the argument set out above.

Data from each laboratory were analysed separately, for several reasons:

1. The task employed in one laboratory [Dataset 1—TC] was the deceptive box task (Perner, Leekam, & Wimmer, 1987) and in the other [Dataset 2—TR, WC] it was the location change task (Wimmer & Perner, 1983).
2. The two laboratories are located in different centres with different socio-economic characteristics (London and Brighton, UK). Systematic biases in the samples from the two laboratories might confound the results in any attempt to combine the two datasets.
3. In one laboratory [Dataset 1], due to differences in the use and scoring of memory and control questions across studies, it was possible only to record a pass/fail, with the criterion adopted that in order to pass the child had to answer naming and reality control questions correctly. In the other laboratory [Dataset 2] several other measures were available which might help to establish whether any female advantage on false belief tasks was independent of language ability. First, a direct measure of verbal ability was also available on approximately one half of cases. Second, data on three memory control questions was recorded separately. Memory questions are of similar verbal complexity as the test question so that if the gender advantage is linguistically mediated we would expect girls to show a substantial advantage on memory questions as well as the test question.

In order to be able to compare gender effects independently in the two datasets but also to examine any systematic age by gender interaction across the two datasets, gender effects on false belief performance were studied independently in each dataset split into age quartiles. If an advantage for girls was to be found we would expect it at one agepoint only, since the advantages offered by the differential socialisation effects outlined above would be transient—at least as measured by a single false belief task. Boys would be expected to catch up once they had accrued the necessary social experience and any continuing gender effects in theory of mind ability would be masked by ceiling effects when measured by a single false belief task.

Method

Dataset 1

(i) Participants. 375 children (183 girls, 192 boys) aged between 2.33 years and 6.25 years (M = 4.24 years, SD = .82 years) from a number of playgroups, nurseries and primary schools in the London area participated. Systematic data on ethnicity and social background was not collected in all studies but the samples were approximately 60% Caucasian, 20% Afro-Caribbean and 20% Asian, and were predominantly from lower middle-class and working class areas of inner and outer London.

For analysis the sample subdivided into age quartiles as follows: 1st quartile: 93 children (46 girls, 47 boys), mean age 3.20 years (SD = .29, range 2.33 to 3.50); 2nd quartile: 95 children (45 girls, 50 boys), mean age 3.96 years (SD = .22, range 3.58 to 4.25); 3rd quartile: 86 children (39 girls, 47 boys), mean age 4.50 years (SD = .13, range 4.33 to 4.67); 4th quartile: 101 children (53 girls, 48 boys), mean age 5.25 years (SD = .42, range 4.75 to 6.25).

(ii) Deceptive box task. Children received a deceptive box false belief task, similar to that developed by Perner et al. (1987)[1]. The child was shown a familiar cereal carton and asked the *Naming Question*: 'What do you think is inside this?' All the children said 'cereal' or named the cereal brand. The carton was then opened and the child was

shown it really contained a mirror. The carton was then closed again, and the child was asked the *Reality Question*: 'Now what do you think is inside this carton?' Most children said 'a mirror', confirming that their belief had now changed. If the child was unable to name the mirror the experimenter informed the child that there was a mirror in the carton and asked them to name the contents again. The child was then asked the *Belief Question*: 'When I first showed you the carton, before we opened it, what did you think was inside?' If the child did not answer, the question was repeated but took the form of 'When I first showed you the carton, before we opened it, did you think there was a mirror or cereal inside?' (order of mirror and cereal counterbalanced across children). In order to pass the task the child had to answer the Belief Question and both Naming and Reality Control Questions correctly.

Dataset 2

(iii) Participants. 1093 children (558 girls, 535 boys) aged between 2.25 years and 6.17 years (M = 3.92 years, SD = .79 years) from a number of playgroups, nurseries and primary schools in the Brighton area participated. Systematic data on ethnicity and social background was not collected in all studies but the large majority of children were Caucasian, and were from middle-class and lower middle class suburban backgrounds.

For analysis the sample subdivided into age quartiles as follows: 1st quartile: 242 children (128 girls, 114 boys), mean age 2.97 years (SD = .24, range 2.25 to 3.32); 2nd quartile: 266 children (131 girls, 135 boys), mean age 3.52 years (SD = .12, range 3.33 to 3.72); 3rd quartile: 299 children (147 girls, 152 boys), mean age 4.04 years (SD = .18, range 3.75 to 4.33); 4th quartile: 286 children (143 girls, 143 boys), mean age 4.97 years (SD = .50, range 4.34 to 6.17).

Location change task:
Children received a location change false belief task, similar in structure to that developed by Wimmer and Perner (1983)[2]. The false belief story was enacted with dolls in two model rooms. In the mouse story, Sam placed his cheese by the blue box and then retreated through the nearby hole behind the wall and out of sight. In Sam's absence, a second mouse, Katy, moved the cheese to her red box. Before Sam returned children were asked the *Belief Question*: 'Where will Sam look for his cheese?' and three *Memory Questions*: *Beginning* 'Where did Sam put his cheese in the beginning?', *Reality* 'Where is the cheese now?', *See* 'Did Sam see Katy move the cheese?'

Data on answers to the Belief Question were available for all 1093 children, and data on the three memory questions were available for 680 children (349 girls, 331 boys)[3]. In addition, verbal mental age data from the British Picture Vocabulary Scale (BPVS; Dunn et al., 1982) were available for 519 children (259 girls, 260 boys).

Results

Dataset 1

The number and percentage of girls and boys passing the deceptive box false belief task at each age quartile are shown in Table 1. In order to test for the predicted age-specific differences separate sequential logistic regressions were conducted on the

Table 1. The Number and Percentage (in Brackets) of Girls and Boys in Each Dataset Broken Down by Age Quartiles Passing the False Belief Tasks

	Girls		Boys	
Age quartile	Fail N (%)	Pass N (%)	Fail N (%)	Pass N (%)
Dataset 1				
1st quartile	38 (82.6%)	8 (17.4%)	39 (83.0%)	8 (17.0%)
2nd quartile	19 (42.2%)	26 (57.8%)	32 (64.0%)	18 (36.0%)
3rd quartile	11 (28.2%)	28 (71.8%)	21 (44.7%)	26 (55.3%)
4th quartile	15 (28.3%)	38 (71.7%)	12 (25.0%)	36 (75.0%)
Dataset 2				
1st quartile	107 (83.6%)	21 (16.4%)	103 (90.4%)	11 (9.6%)
2nd quartile	85 (64.9%)	46 (35.1%)	99 (73.3%)	36 (26.7%)
3rd quartile	72 (49.0%)	75 (51.0%)	71 (46.7%)	81 (53.3%)
4th quartile	38 (26.6%)	105 (73.4%)	31 (21.7%)	112 (78.3%)

sample split by cumulative age quartile, with age entered into the prediction equation first. Entering age first ensured that differences in mean age between boys and girls were not responsible for gender effects. In the 1st (youngest) quartile of the sample neither age nor gender significantly predicted false belief performance (X^2 (1, N = 93) = .018; X^2 (1, N = 93) = .001, respectively, both $p > .10$). In the 1st plus 2nd quartiles combined age significantly predicted false belief performance and there was a non-significant trend to gender predicting false belief performance (X^2 (1, N = 188) = 15.0, $p < .001$; X^2 (1, N = 188) = 3.35, $p = .07$, respectively). In the 1st, 2nd and 3rd quartiles combined age and gender significantly predicted false belief performance (X^2 (1, N = 274) = 37.6, $p < .001$; X^2 (1, N = 274) = 5.94, $p < .05$, respectively). In the whole sample age significantly predicted false belief performance and there was a non-significant trend to gender predicting false belief performance (X^2 (1, N = 375) = 75.5, $p < .001$; X^2 (1, N = 375) = 3.51, $p = .06$, respectively).

Regression coefficients, odds ratios and 95% confidence intervals for odds ratios for the predictor variables age and gender for Dataset 1 broken down by cumulative age quartiles are shown in Table 2. The odds ratio indicates the relative size of the effect of age and gender on false belief performance. That is, it shows the change in the relative odds of a subject passing or failing the false belief task when the predictor variable of interest changes by a unit of one (in this case one year for age, and being a boy vs. a girl for gender). Comparison of the size of the odds ratios of the two predictor variables indicates their relative strength, although this is more conceptually difficult when one variable is continuous and the other is categorical, as in the present case. In the 1st, 2nd and 3rd age quartiles combined where the effect of both age and gender significantly predicted false belief performance, the relative increase in likelihood of a subject passing the false belief task was approximately 9 times higher for each year of age than for being a girl rather than a boy.

Table 2. Regression Coefficients, Odds Ratios and 95% Confidence Intervals for Odds Ratios for Age in Years and Gender in Each Dataset Broken by Cumulative Age Quartiles

Age quartiles	Predictor	B	Odds ratio	95% CI
Dataset 1				
1st quartile	Age	−.13	.88	.14 to 5.61
	Gender	−.02	.98	.33 to 2.89
1st + 2nd quartile	Age	1.50	4.50	2.05 to 9.90
	Gender	−.60	.55	.29 to 1.05
1st + 2nd + 3rd quartile	Age	1.51	4.51	2.70 to 7.52
	Gender	−.65	.52	.31 to .88
1st + 2nd + 3rd + 4th quartile	Age	1.28	3.58	2.58 to 4.96
	Gender	−.43	.65	.41 to 1.02
Dataset 2				
1st quartile	Age	1.68	5.37	.83 to 34.9
	Gender	−.68	.51	.23 to 1.11
1st + 2nd quartile	Age	2.00	7.31	3.32 to 16.1
	Gender	−.48	.62	.40 to .96
1st + 2nd + 3rd quartile	Age	1.96	7.07	4.80 to 10.4
	Gender	−.23	.79	.58 to 1.09
1st + 2nd + 3rd + 4th quartile	Age	1.60	4.96	3.97 to 6.19
	Gender	−.09	.91	.69 to 1.20

Dataset 2

The number and percentage of girls and boys passing the location change false belief task at each age quartile are shown in Table 1. Again, separate sequential logistic regressions were conducted on the sample divided by cumulative age quartiles, with age entered into the prediction equation first. In the 1st quartile neither age nor gender predicted false belief performance, although both showed a non-significant trend towards doing so (X^2 (1, N = 242) = 2.94, p = .09; X^2 (1, N = 242) = 2.98, p = .08, respectively). In the 1st and 2nd quartiles combined both age and gender predicted false belief performance (X^2 (1, N = 508) = 27.7, p < .001; X^2 (1, N = 508) = 4.71, p < .05, respectively). In the 1st, 2nd and 3rd quartiles combined and in the whole sample age but not gender significantly predicted false belief performance (X^2 (1, N = 807) = 118.8, p < .001; X^2 (1, N = 807) = 2.03, p > .10; and X^2 (1, N = 1093) = 288.8, p < .001; X^2 (1, N = 1093) = .43, p > .10, respectively). Odds ratios and other relevant statistics are shown in Table 2. In the 1st and 2nd age quartiles combined where the effect of both age and gender significantly predicted false belief performance, the relative increase in likelihood of a subject passing the false belief task was approximately 12 times higher for each year of age than for being a girl rather than a boy.

For a subset of this dataset we also had data on girls' and boys' performance on the three memory questions and the BPVS. These effects were explored for the older and younger half of the sample, split by median age. On the memory questions there was no difference between girls' and boys' performance in either the younger (ANOVA; $F(1, 298) = .06$, p = n.s.; girls: $M = 2.33$, $SD = .89$; boys: $M = 2.30$, $SD = .92$) or the older half of the sample (ANOVA; $F(1, 379) = .009$, p = n.s.; girls: $M = 2.36$, $SD = .62$; boys: $M = 2.35$, $SD = .67$). Similarly, BPVS scores showed no significant advantage for girls over boys in either the younger (ANOVA; $F(1, 164) = .15$, p = n.s.; girls: $M = 41.5$ months, $SD = 10.0$ months; boys: $M = 40.8$ months, $SD = 11.1$ months) or the older half of the sample (ANOVA; $F(1, 351) = 1.44$, p = n.s.; girls: $M = 57.4$ months, $SD = 15.5$ months; boys: $M = 55.6$ months, $SD = 13.6$ months).

Discussion

A significant but weak female advantage was found for false belief task performance in the 1st, 2nd and 3rd cumulative age quartiles in Dataset 1 and a similar significant but weak effect was found in the 1st plus 2nd age quartiles combined in Dataset 2. Over both datasets, although girls did better than boys, the gender advantage was weak and substantially weaker than the effect for age.

Two potentially complimentary explanations for this weak, age-specific theory of mind advantage are apparent. First, previously identified gender differences in the amount of supportive and emotion talk that mothers enter into with their offspring (Cervantes & Callanan, 1998; Dunn et al., 1987; Leaper et al., 1998), and older siblings with younger siblings (Brown & Dunn, 1996; Brown et al., 1996), both favour girls over boys. This difference in social milieu and social experience may 'boost' girls to acquire theory of mind understanding, on average, slightly earlier than boys. One unknown is exactly how facilitation of false belief takes place. It is possible that even if mothers do enhance false belief understanding, they do so without directly instructing about beliefs. Perhaps the increased emotion talk with girls in non-disciplinary situations (Brown et al., 1996; Cervantes & Callanan, 1998; Dunn et al., 1987; Hughes et al., 1999), and in disciplinary situations (Ruffman et al., 1999), is sufficient in getting children to think more carefully about social relations in general. When measured by a single yardstick, as in this study, the advantage is merely temporary and boys eventually acquire a sufficient quality and quantity of social experience in order to understand first order false beliefs so that the gender difference disappears for older children.

The second explanation is that there is an innate female advantage in mindreading abilities (Baron-Cohen & Hammer, 1997; Skuse, 1998). These accounts draw on data from both typically-developing adults and children, and from clinical groups who are considered to show an 'extreme' male cognitive style—individuals with autism and individuals with Turner's syndrome who are monosomic with a maternally-acquired X chromosome. Recent research has found a high heritability in theory of mind in typically-developing pre-school, same-sex twin pairs, independent of verbal ability (Hughes & Cutting, 1999).

The socialisation and nativist accounts may indeed be complementary in that a genetic predisposition to engage in or understand certain types of material is likely to influence a child's environment (e.g. how much their mother or older siblings talks to them about such matters) in ways that may increase the innately-specified advantage. Such gene-environment interactions are becoming better understood in child development (Plomin & Rutter, 1998), and gender differences in the development of social and emotional understanding and any possible links to neuropsychiatric disorders will be an important area for future study. However, the present findings cannot inform this debate in terms of the relative contribution of, and role of gene-environment interaction in, genetic and environmental influences on gender differences in mental state understanding (see Hughes & Cutting, 1999, for a discussion of these issues).

Another possible explanation for the gender effect is that the age-specific false belief advantage for girls is mediated by general language ability. For instance, it has been shown that language ability is related to theory of mind ability (Astington & Jenkins, 1999; de Villiers, 2000; Happé, 1995; Jenkins & Astington, 1996). In a subsample of Dataset 2 we found that girls were not better on either the BPVS, which measures receptive vocabulary, or on the memory questions that are of a similar verbal complexity to the false belief test question. However, a recent view is that syntax is most closely associated with theory of mind ability in both typical (de Villiers, 2000) and atypical groups (Tager-Flusberg, 2000) and no direct measure of syntactic competence was available in the present datasets. A variant on the language hypothesis is that mediation is more evident in the early phase of language acquisition in comparison to the latter phase (e.g. Bornstein & Haynes, 1998; Huttenlocher et al., 1991; Hyde & Linn, 1988). We cannot rule out this explanation with our data but future studies could examine this by including more extensive tests of language abilities that are related to belief understanding (Astington & Jenkins, 1999; de Villiers, 2000).

The strength of this study is in the very large numbers of children tested, overcoming the insufficient statistical power in previous studies to identify a weak effect that is apparent in a limited time window using a single task. With such large numbers and with similar patterns across two independent laboratories we can be relatively certain that our effects are representative of genuine gender differences. Yet the necessity to test such large numbers to get the effect, and the fact that the effect for gender is weak is equally important. In this study the analysis of gender was conducted post-hoc on existing datasets. Future research into gender effects on theory of mind development should adopt prospective designs to test in more detail the a priori predictions that follow from the psychocultural (Maccoby, 2000), psychosocial (Hughes & Dunn, 1998) and nativist (Baron-Cohen & Hammer, 1997) accounts. In addition this research will need to take account of the potentially mediating role of factors such as language development and parenting style.

What should one make of a relatively weak female advantage on theory of mind tasks? On the positive side we note that the advantage, although weak, seems to persist throughout much of development in that researchers have obtained such findings with preschool children, pre-adolescent children, and young adults. Second, the advantage has been linked to female advantages in social competence in pre-adolescents and adolescents (Bosacki & Astington, 1999; Mathews & Keating, 1995), and to the largely male condition of autism (Baron-Cohen & Hammer, 1997; Skuse, 2000). Thus, the advantage, even if weak, might be of some consequence for both males and females. Yet on the negative side, we think that many and possibly even most individuals will not show a gender-typed pattern of performance. Gender effects are weak and might perhaps be best construed as general tendencies with exceptions being the norm. Other psychosocial factors such as family environment (Perner et al., 1994), parenting style (Hughes et al., 1999; Ruffman et al., 1999) and interaction with siblings, other family members and peers (Hughes & Dunn, 1998; Lewis et al., 1996) may be more important in determining the rate and range of mentalising ability that individual children acquire.

References

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

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

Baron-Cohen, S. & Hammer, J. (1997). Is autism an extreme form of the 'male brain'? *Advances in Infancy Research*, *11*, 193–217.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813–822.

Bornstein, M. H. & Haynes, O. M. (1998). Vocabulary competence in early childhood: measurement, latent construct and predictive validity. *Child Development*, *69*, 654–671.

Bosacki, S. & Astington, J. W. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development*, *8*, 237–255.

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

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.

Cervantes, C. A. & Callanan, M. A. (1998). Labels and explanations in mother-child emotion talk: age and gender differentiation. *Developmental Psychology*, *34*, 88–98.

Cohen, D. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: LEA.

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

De Villiers, J. (2000). Language and theory of mind: What are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.) *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience (2nd Edn.)*, pp. 83–123. Oxford: Oxford University Press.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: individual differences and their antecedents. *Child Development*, *62*, 1352–1366.

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology*, *23*, 132–139.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintillie, D. (1982). *British Picture Vocabulary Scale*. London: NFER-Nelson.

Fonagy, P., Redfern, S., & Charman, T. (1997). Individual differences in theory of mind acquisition: the role of attachment security. *British Journal of Developmental Psychology*, *15*, 51–61.

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.

Happé, F. G. E., Winner, E., & Brownell, H. (1998). Linking theory of mind and central coherence bias in autism and in the general population. *Developmental Psychology*, *34*, 358–362.

Holmes, H. A., Black, C., & Miller, S. A. (1996). A cross-task comparison of false belief understanding in a Head Start population. *Journal of Experimental Child Psychology*, *63*, 263–285.

Hsieh, F. Y., Bloch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, *17*, 1623–1634.

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

Hughes, C., Deater-Deckard, K., & Cutting, A. L. (1999). 'Speak roughly to your little boy'? Sex differences in the relations between parenting and preschooler's understanding of mind. *Social Development*, *8*, 143–160.

Hughes, C. & Cutting, A. L. (1999). Nature, nurture and individual differences in early understanding of mind. *Psychological Science*, *10*, 429–432.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991) Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, *27*, 236–248.

Hyde, J. S. & Linn, M. C. (1988) Gender differences in verbal ability: a meta-analysis. *Psychological Bulletin*, *104*, 53–69.

Jarrold, C., Butler, D. W., Cottington, E. M., & Jimenez, F. (2000). Linking theory of mind and central coherence bias in autism and in the general population. *Developmental Psychology*, *36*, 126–138.

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.

Leaper, C., Anderson, K. J., & Sanders, P. (1998). Moderators of gender effects on parents' talk to their children: a meta-analysis. *Developmental Psychology*, *34*, 3–27.

Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki Kassotaki, K., & Berridge, D. M. (1996). Social influences on false belief access: specific sibling influences or general apprenticeship? *Child Development*, *67*, 2930–2947.

Maccoby, E. E. (2000). Parenting and its effects on children: On reading and misreading behavior genetics. *Annual Review of Psychology*, *51*, 1–27.

Mathews, D. & Keating, D. (1995). Domain specificity and habits of mind: An investigation of patterns of high-level development. *Journal of Early Adolescence*, *15*, 319–343.

Meins, E., Fernyhough, C., Russell, J., & Clark-Carter, D. (1998). Security of attachment as a predictor of symbolic and mentalising abilities: a longitudinal study. *Social Development*, *7*, 1–24.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development*, *65*, 1228–1238.

Plomin, R. & Rutter, M. (1998). Child development, molecular genetics, and what to do with genes once they are found. *Child Development*, *69*, 1223–1242.

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, *34*, 161–174.

Ruffman, T., Perner, J., & Parkin, L. (1999). How parenting style affects false belief development. *Social Development*, *8*, 395–411.

Skuse, D. H. (2000). Imprinting, the X-chromosome and the male brain: Explaining sex differences in the liability to autism. *Paediatric Research*, *47*, 9–16.

Tager-Flusberg, H. (2000). Language and understanding minds: Connections in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.) *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience (2nd Edn.)*, pp. 124–149. Oxford: Oxford University Press.

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.

Notes

1. Across the pooled dataset the contents and the container varied (a Smarties box containing a pencil, an eggbox containing a tomato, a Bandaid box containing an eraser) but the procedure and the test questions were invariant.
2. Across the pooled dataset the characters and the containers used, and the contents which are moved, varied but all had an invariant structure in which one character placed something in one of the two containers and was absent when the second story character moved the contents to the second container.
3. The proportion of subjects in the older and younger halves of the sample for whom memory question responses were recorded varied, as did the proportion who completed the BPVS. The same age quartile boundaries are used in order that children remain in the same age group across the 2 sets of analyses.