Research Article

Lost in Translation? Comparing British, Japanese, and Italian Children's Theory-of-Mind Performance

Claire Hughes,[1] Rory T. Devine,[1] Rosie Ensor,[1] Masuo Koyasu,[2] Ai Mizokawa,[2] and Serena Lecce[3]

[1] *Centre for Family Research, University of Cambridge, Free School Lane, Cambridge CB2 3RQ, UK*
[2] *Graduate School of Education, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto-shi, Kyoto 606-8501, Japan*
[3] *Department of Humanities, University of Pavia, Piazza Botta, 6-27100 Pavia, Italy*

Correspondence should be addressed to Claire Hughes; ch288@cam.ac.uk

Findings from cross-cultural theory-of-mind studies highlight potential measurement effects and both general (e.g., East-West) and specific (e.g., pedagogical experiences) cultural contrasts. We compared theory-of-mind scores for children from UK and Italy (two Western countries that differ in age of school entry) and Japan (a Far-Eastern country in which children, like their Italian counterparts, start school later than British children). Confirmatory factor analysis was applied to data from 268 age-gender- and verbal ability-matched 5- to 6-year olds. Key findings were that (i) all 8 indicators loaded onto a single latent factor; and (ii) this latent factor explained significant variance in each group, with just one indicator showing differential item functioning. Supporting the importance of pedagogical experiences, British children outperformed both their Italian and Japanese counterparts.

1. Introduction

Individual differences in the rate at which children acquire a theory of mind appear important for children's success at school [1, 2] and for relationships with friends (e.g., [3]) and peers [4, 5]. However, the lion's share of research on individual differences in children's understanding of mind has focused on the origins of individual differences. As documented in recent reviews (e.g., [6, 7]) much of this research has concerned structural family factors, such as number of siblings [8] or overall family size [9], or more qualitative family factors, such as frequency of maternal talk about mental states (e.g., [10–12]) or cooperative interactions with siblings [13, 14].

Beyond the family, other studies have shown that variation in conversations about mental states with friends [15] and children's social acceptance by their peer group [4] also predict individual differences in children's performance on tests of theory of mind. Alongside this research on extrafamilial social influences is a marked expansion in the scope of

theory-of-mind research, such that accounts that emphasize universality (e.g., [16, 17]) have been challenged on several fronts. Specifically, meta-analytic findings indicate that Asian children lag significantly behind American and British children on false-belief tasks [18]. In addition, children from different cultures appear to vary not only in the rate but also in the order in which they achieve distinct milestones within theory of mind. For example, while children from individualistic cultures typically acquire an understanding of the subjective nature of belief before they appreciate constraints on knowledge, children from collectivistic cultures, such as China or Iran, typically understand knowledge before they understand belief [19, 20].

Theoretically, findings from cross-cultural studies provide not only a rigorous test of the generalizability of theories that emerge from one particular culture but also a means of testing commonly held assumptions regarding the structure of children's theories of mind. In particular, do distinct measures tap into the same latent ability even when administered in different languages to children from different

cultures? Existing work on East-West differences in children's understanding of mind is largely restricted to comparisons of first-order false-belief understanding in children from China and USA (e.g., [18, 21]), with converging support from studies that examine false-belief understanding in children from other Confucian countries, including Korea [22, 23] and Japan [24, 25]. Collectively, these findings lead to the prediction that children from collectivist Confucian cultures (such as China, Korea, and Japan) are likely to lag behind children from individualistic Western cultures in their understanding of subjective mental states [22, 26]. Challenging this simple East-West contrast, however, are findings that suggest significant variation *within* each of these cultural settings. First, meta-analytic data reveal that children from Hong Kong lag behind their peers in mainland China [18], despite leading much more westernized lives. One possible explanation for this contrast hinges on differences in children's linguistic environments: for economic reasons, a very large proportion of children in Hong Kong are cared for by adults from outside the family who may share their bedroom but typically do not speak Cantonese (e.g., [27]). Second, Lecce et al. [2] reported a similarly surprising contrast between two Western groups: children in UK and Italy (matched for age, gender, verbal ability, and maternal education). In discussing this contrast, these authors noted that Italian children begin formal schooling at age 6, at least a year later than their British counterparts. This raises the possibility that entry to school is at least as important as the contact with close kin that is such a feature of Mediterranean life [9]. Specifically, formal schooling offers both increased contact with peers and exposure to pedagogical situations that encourage reflective self-awareness and so may be particularly important to theory-of-mind development [28].

In the current study, we built on existing collaborative relations to recruit samples in UK, Italy, and Japan. Existing research into the East-West contrast in theory of mind has been almost exclusively focused on the performance of 3- to 5-year olds on standard first-order false-belief tasks. To test the developmental and methodological generalizability of existing findings, we selected school-aged children (5- to 6-year-olds) and included a variety of theory-of-mind tasks in our test battery. To our knowledge, the current study is the first cross-cultural to include not only standard first-order false-belief tasks but also tasks that tap into children's later-emerging abilities to infer emotion from false belief [29] and to understand mistaken beliefs about others' beliefs, that is, second-order false-beliefs [30]. A further limitation of previous cross-cultural studies is that these have not typically assessed participants' verbal ability (e.g., [16, 17]) or matched samples on both age and verbal ability (e.g., [11]). Given that there is a widely reported robust relation between verbal ability and false-belief understanding (e.g., [31]), this omission is surprising. In the current study we therefore oversampled within each country to ensure that it would be possible to match groups for age, gender, and verbal ability.

Thus, by administering a wide variety of theory-of-mind tasks to carefully matched groups of school-aged children, we aimed both to extend the conceptual and developmental scope of existing cross-cultural research and to improve the methodological rigour of this research. In particular, adopting a battery approach allowed us to apply statistical methods to test the extent to which cultural differences simply reflect contrasts in the measurement properties of tasks used in different languages with different cultural groups. Group differences can arise in cross-cultural research for a number of reasons that are unrelated to differences in underlying abilities; these include differing definitions and meanings of a concept, inappropriate translations, and differing response styles, reflecting differences in social norms [32, 33]. Nevertheless, group comparisons have traditionally used statistical procedures (e.g., t-tests) that assume that test items function in a similar way for all participants regardless of group membership [32]. This assumption of measurement invariance (i.e., equivalent empirical relations in different groups between test items and the latent construct) means that group comparisons can yield spurious and misleading results.

Recent advances in the measurement and analysis of children's understanding of mind are therefore useful in strengthening the methodological rigor of cross-cultural research. First, by testing the across-group equivalence of the structure and scale of the latent construct, multiple-groups confirmatory factor analysis (CFA) [34] provides a statistical means of assessing measurement invariance. Thus, our first aim was to assess whether commonly used tests of children's theory of mind exhibited measurement invariance across three different cultures. To our knowledge, this is the first cross-cultural study of children's theory-of-mind development to undertake a direct assessment of measurement invariance.

In summary, our study had two primary aims. Our first aim was to apply multiple-groups CFA to examine whether a battery of theory-of-mind tasks would exhibit the same measurement properties (i.e., equal form, equal factor loadings, and equal factor variance) in children from UK, Italy, and Japan. Our second aim was to test two competing hypotheses. The first, based on the potential importance of contrasts between Western and Confucian cultures, was that both British and Italian children would outperform their Japanese counterparts on the theory-of-mind task battery. The second, based on the potential importance of pedagogical experiences for children's awareness of others' minds, was that children from UK, who enter school a year earlier than children in either Italy or Japan, would outperform these two groups (who would perform similarly to each other) on the theory-of-mind task battery.

2. Method

2.1. Participants. Recruitment in UK, Italy, and Japan took place via primary schools in three similarly historic cities: Cambridge, Pavia, and Kyoto (the Universities of Cambridge and Pavia were founded in 1209 and 1311 (resp.,) and Kyoto was the capital of Japan from 794 to 1868). In size, Cambridge and Pavia are similar in terms of both surface area ($115\,km^2$ and $62\,km^2$, resp.,) and population size (123,900 and 71,000, resp.,) but much smaller than Kyoto ($828\,km^2$,

Table 1: Proportion of participants passing theory-of-mind tasks.

Indicator	Total sample	Gender differences			National group differences			
		Males	Females	χ^2	UK	Japan	Italy	χ^2
First-order false belief (nice surprise)	.78	.78	.77	.05	.81	.84	.68	7.77
First-order false Belief (nasty surprise)	.88	.87	.88	.05	.93	.85	.84	4.03
First-order false belief (chocolate)	.70	.70	.70	.01	.89	.73	.49	34.80[*]
First-order false belief (puppy)	.56	.59	.54	.57	.88	.53	.28	66.32[*]
Infer emotion from false belief (nice)	.63	.65	.61	.63	.62	.65	.62	.17
Infer emotion from false belief (nasty)	.71	.70	.71	.05	.78	.57	.77	11.86[*]
Second-order false belief (chocolate)	.34	.32	.35	.19	.43	.31	.27	6.09
Second-order false belief (puppy)	.24	.26	.22	.62	.29	.22	.20	2.25

Note. [*] $P < .0062$ (Bonferroni adjustment to compensate for multiple tests).

population = 1.4 million). On average, residents in all three cities are relatively prosperous: Cambridge residents have a median annual income of $45,000 as compared with $29,040 for Pavia and $41,850.68 for Kyoto [35–37].

We initially recruited 118 children from UK, 106 children from Japan and 123 children from Italy. We then matched participants on the basis of age, gender, and performance on a measure of receptive vocabulary. The final sample therefore consisted of 268 children (UK: N = 45 males, 45 females; Japan: 43 males, 45 females; Italy: 45 males, 45 females), with an average age of 6.05 years (SD = .29; UK: M Age = 6.03, SD = .32; Japan: M Age = 6.06, SD = .30; Italy: M Age = 6.06, SD = .24). Note there were fewer males overall in the original Japanese sample of 106 children and so we could not obtain precisely matched samples. Table 1 presents sample characteristics and descriptive statistics for each national group.

2.2. Procedures and Measures. Each of the study tasks was translated into Italian and Japanese and then backtranslated (by native speakers of each language who were also fluent in English) in order to check the accuracy of these translations. Most of the tasks used had been translated into Italian for a previous study, and the actual wording (in both English and Italian) is presented as an appendix in the paper by [38]. Tasks were administered individually in a quiet room at school by the authors and/or experienced and trained researchers, using ethnically neutral props (e.g., toy animals, cartoon drawings) and a Latin-square design to counterbalance the order of presentation within the theory-of-mind task battery and between this task battery and the verbal ability test. The same puppets, story books, and materials were used in each site.

2.2.1. Theory of Mind. We assessed children's understanding of mental states using four separate tasks which have previously shown good psychometric properties [3, 39].

The first two tasks were enacted using puppets and featured stories about either a nice or a nasty surprise [29]. In one story ("The Nasty Surprise") a character (Leo the Lion) received a nasty surprise when another character (Croc the Crocodile) played a trick on him by substituting the contents of a can of his favorite drink (cola) with a drink he disliked (juice). In another story ("The Nice Surprise") a character received a nice surprise when another character (Freddie the Frog) removed his least favorite snack (a pear) from his lunchbox and replaced it with his favorite snack (an apple). Previous research has shown that, for typically developing children at least, performance on these two tasks is similar and highly correlated [39].

In both stories, the examiners asked participants two forced-choice comprehension questions to establish they understood the characters' desires (e.g., "How does Monty feel when he gets a pear? Does he feel happy or not happy?"). Next both stories contained a question that required children to predict a character's false belief about the contents of a container (either the drink carton or the lunch box) (e.g., "What does Monty think is in the box, an apple or a pear?") and a forced-choice control question about the actual contents of the container (e.g., "What is in the box really, an apple or a pear?"). The participants were only credited with passing the false-belief prediction question if they passed both the test and control question.

Next, the examiners asked the participants to make an emotion inference based upon both desire and belief information [29]. Specifically, the examiners asked the participants a forced-choice question about the character's emotional state about the contents of the container (either the lunch box or the carton) before opening it (e.g., "How does he feel before he looks in the box? Is he happy or not happy?"). Following this, participants were asked an explanation question (e.g., "Why is he happy?"). In addition to these two questions, participants were also required to pass two further comprehension questions about the character's emotional state after learning the actual contents of the container. Participants were only credited with being able to infer a character's emotions based on beliefs and desires if they passed both (a) the relevant false-belief prediction question and (b) the comprehension questions about the character's emotional state after learning the true contents of the container. We derived two separate scores from participants' performance on each task: (a) the ability to predict a first-order false belief (0-1) and (b) the ability to infer and explain a character's emotion based on his/her false-beliefs (0-1). Thus, from these two stories we created four separate indicators: First-Order

False Belief (Nice Surprise), First-Order False Belief (Nasty Surprise), Infer Emotion from False Belief (Nice Surprise), and Infer Emotion from False Belief (Nasty Surprise).

The remaining two tasks featured picture book stories designed to measure children's understanding of second-order false beliefs [40]. In the first story ("The Puppy Story"), Peter's mother buys him a puppy for his birthday and wants to surprise him. To do so, she tells Peter that she will not get him a puppy. Unbeknown to Peter's mother, he discovers the puppy while looking for his bicycle in the shed. In this task, participants first had to answer a forced-choice question about Peter's initial false belief about his birthday present ("What did Peter think he was getting for his birthday?") and a corresponding question about what Peter is really getting for his birthday ("What was his mother giving him really?"). Four further questions were asked to assess second-order false-belief understanding, that is, the ability to make inferences about a character's beliefs about another character's beliefs [41]. In the story, Peter's grandmother telephones his mother and asks her what Peter thinks he is getting for his birthday. Children were asked two further test questions ("What does Mum say to Granny?" and "Why did she say that?") as well as two further control questions ("Did Mum see Peter finding the present?" and "What has Mum really got Peter for his birthday?").

In the second story ("The Chocolate Story"), Mary and John are given some chocolate to share. John hides the chocolate in a tin but lies to Mary about its location telling her it is in the refrigerator. Later, unbeknown to John, Mary sees him taking some of the chocolate out of the tin. As in the previous story, the examiners asked participants to answer a forced-choice question about Mary's initial false belief about the location of the chocolate ("Where does Mary think the chocolate is?") and a corresponding question about the actual location of the chocolate ("Where did John put the chocolate really?"). After Mary and John have finished playing, their mother tells Mary that she can come in to get the chocolate. Participants were then asked a second-order false-belief attribution question ("Where does John think that Mary will look for the chocolate?") and an explanation question ("Why does he think that?"). These questions were followed by two further control questions to assess children's comprehension of the story ("Did John see Mary looking through the kitchen window?" and "Where did John put the chocolate really?").

From each of these two stories, we derived two separate scores: (a) the ability to predict a character's first-order false belief (0-1) and (b) the ability to infer and explain a character's second-order false belief (0-1). Participants were only credited with correctly passing the first-order false-belief question if they passed both the test question and reality control question. Participants were only credited with passing the second-order false-belief question if they passed both the inference and explanation questions as well as the two comprehension questions. Thus across these two stories, we obtained four further binary indictors: First-Order False Belief (Chocolate Story), First-Order False Belief (Puppy Story), Second-Order False Belief (Chocolate Story), and Second-Order False Belief (Puppy Story).

2.2.2. Verbal Ability. Verbal ability (VA) was measured using the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) receptive vocabulary test [42] or a language-appropriate translation [43, 44]. Participants were required to point to one of four pictures that matched a word read aloud by the examiner. Testing was discontinued after five consecutive incorrect responses. The total number of correct items were summed together to create a score for each participant giving a possible range of 0 to 38.

3. Results

National groups did not differ by age, $F(2, 265) = .27$, $P = .76$, or gender, $\chi^2(2) = .03$, $P = .99$, or receptive vocabulary scores, $F(2, 265) = .89$, $P = .89$. The mean receptive vocabulary score was 29.19, SD = 3.21 (UK: M = 29.47, SD = 3.58, Japan: M = 29.27, SD = 2.96, Italy: M = 28.84, SD = 3.05). There were no gender differences in age, $t(266) = .65$, $P = .52$, or receptive vocabulary scores, $t(266) = .45$, $P = .65$.

During the matching procedure we excluded those children with missing data on key matching variables (i.e., age, gender, and receptive vocabulary). We did not exclude children who failed control questions on the theory-of-mind tasks. Instead, children who failed the control questions were allotted a score of zero on the corresponding test question. Table 1 shows the proportion of children passing each of the theory-of-mind indicators within the total sample, the male sample, the female sample, and each national group. Table 2 shows tetrachoric correlations between individual indicators (these describe the relations between two dichotomous variables when it is assumed that a continuous latent variable underpins test performance [45]). As Table 1 shows, there were no gender differences for any of the theory-of-mind task indicators.

3.1. Analytic Strategy. We used MPlus Version 6 [46] to examine the latent factor structure of the 8 binary (pass/fail) indicators of theory of mind. Given the categorical nature of the data, we used a mean- and variance-adjusted weighted least squares estimator (WLSMV) [34, 47]. This approach assumes that performance on dichotomous categorical indicators is related to a continuous and normally distributed underlying ability. That is, a certain amount of the underlying trait or ability is needed to pass the threshold on each indicator (e.g., from fail to pass) ([34]; Kline, 2011). Thus estimates are based on the tetrachoric correlation matrix for the binary indicators. We evaluated model fit using two recommended criteria: comparative fit index (CFI) >.90 and tucker lewis index (TLI) >.90 (e.g., [34]). The difference in fit between nonnested models was assessed using the Akaike information criterion (AIC) with preference given to the model with the lowest AIC value [34]. Since the difference in χ^2 values for nested models estimated using WLSMV does not adhere to the standard χ^2 distribution, the difference in fit between nested models was evaluated using a special formula for the χ^2 difference test available in MPlus [34].

TABLE 2: Tetrachoric correlation matrix for theory-of-mind indicators.

	1	2	3	4	5	6	7	8
(1) First-order false Belief (nice surprise)	—							
(2) First-order false belief (nasty surprise)	.70	—						
(3) First-order false belief (chocolate)	.66	.51	—					
(4) First-order false belief (puppy)	.53	.61	.71	—				
(5) Infer emotion from false belief (nice)	.91	.69	.58	.43	—			
(6) Infer emotion from false belief (nasty)	.35	.70	.22	.33	.35	—		
(7) Second-order false belief (chocolate)	.57	.65	.88	.57	.55	.32	—	
(8) Second-order false belief (puppy)	.54	.55	.64	.87	.56	.28	.70	—

Note: all values are significant, $P < .001$.

3.1.1. Modelling Individual Differences in Theory of Mind. Our first step was to compare the fit of three competing hypothetical models of the structure of individual differences in theory of mind in the sample as a whole. In the first model, we specified a one-factor solution in which each of the 8 binary indicators loaded onto a single latent theory-of-mind factor. All but one of the fit indices suggested that the model provided an adequate fit to the data, $\chi^2(20) = 105.89$, $P < .0001$, CFI = .94, TLI = .92. Modification indices highlighted that the model fit could be improved by permitting measurement error terms to be correlated for the Predict First-Order False Belief indicator and corresponding Infer Emotion based on False Belief indicator for both the Nice Surprise and Nasty Surprise stories. Correlated measurement error arises when two test indicators are related by both the common latent factor and other sources such as shared methods of administration [34]. We adjusted the model accordingly given that the Predict First Order False Belief and Infer Emotion based on False Belief indicators arose from the same task. This modified model provided a good fit to the data, $\chi^2(18) = 42.06$, $P = .001$, CFI = .98, TLI = .98, AIC = 6.06.

Next we specified an alternative model with two correlated latent factors. The first latent factor was comprised of six indicators: four first-order false-belief indicators and two second-order false-belief indicators. The second latent factor consisted of the two emotion-inference indicators. We permitted the measurement terms for the Predict First-Order False Belief and Infer Emotion based on False Belief to correlate within the Nice Surprise and Nasty Surprise stories. This model provided a good fit to the data, $\chi^2(17) = 40.83$, $P = .001$, CFI = .98, TLI = .98, AIC = 6.83.

In the third and final model, we specified a three-factor solution. The first latent factor consisted of four first-order false-belief indicators, the second consisted of two second-order false-belief indicators, and the third consisted of two emotion-inference based on false belief indicators. These three latent factors were permitted to correlate. Consistent with the previous two models, we correlated the measurement error terms for the Predict First-Order False Belief and the corresponding Infer Emotion based on False Belief indicator for the Nice Surprise and Nasty Surprise stories. This third model provided an adequate fit to the data, $\chi^2(15) = 41.25$, $P = .001$, CFI = .98, TLI = .97, AIC = 11.25.

We selected the first model (the one-factor solution) over the second and third model as this model provided the most parsimonious solution to the data. This model exhibited the lowest AIC value and was theoretically the most simple model. The theory-of-mind latent factor explained significant variance in performance on the 8 indicators, unstandardized estimate = 0.56, $P < .001$. Table 3 presents the standardised item loadings for the theory-of-mind latent factor.

3.2. Modelling Theory-of-Mind Performance across Nations. Prior to examining mean performance on the latent theory-of-mind factor in each national group, we used multiple-groups CFA to examine whether the single latent factor solution exhibited measurement invariance across the three groups. We replicated the single latent-factor solution holding the form, factor loadings, and indicator thresholds equal across all three groups. With the exception of one fit index, this model exhibited adequate fit, $\chi^2(66) = 115.14$, $P = .0002$, CFI = .97, TLI = .97. Inspection of the modification indices showed that one indicator was noninvariant, specifically the Predict First-Order False Belief item from the Peter and the Puppy Story. That is, there was a difference in performance on this particular indicator that was unrelated to children's underlying theory-of-mind abilities. The identification of a noninvariant item does not preclude further measurement invariance testing [48]. A model is said to exhibit partial measurement invariance if a selection of items are noninvariant [48, 49]. Some researchers may opt to eliminate noninvariant items. However, this practice can result in incomplete coverage of the construct or the creation of different scales for different groups [33]. The partial measurement invariance approach, in which invariant items are constrained to equality across groups and noninvariant items are free to vary across groups, reduces bias in the model and uses all the data available [33, 34]. This method permits the examination of latent factor mean differences even in the presence of noninvariant items [34].

We therefore freely estimated the thresholds for the noninvariant indicator in UK and Italy. That is, while all other factor loadings and thresholds were constrained, the Predict First-Order False Belief item from the Peter and the Puppy Story was released. The resulting model provided a good fit to the data, $\chi^2(64) = 91.51$, $P = .01$, CFI = .99, TLI = .98. This suggests that the single theory-of-mind latent

TABLE 3: Standardized WLSMV estimates of factor loadings for the theory-of-mind latent factor.

	Total Sample	Multiple-groups CFA		
		UK	Japan	Italy
First-order false belief (nice surprise)	.75	.65	.92	.75
First-order false belief (nasty surprise)	.81	.69	.86	.78
First-order false belief (chocolate)	.82	.98	.76	.86
First-order false belief (puppy)	.84	.71	.82	.97
Infer emotion from false belief (nice)	.68	.63	.92	.65
Infer emotion from false belief (nasty)	.39	.44	.25[+]	.61
Second-order false belief (chocolate)	.78	.75	.75	.87
Second-order false belief (puppy)	.89	.89	.79	.93

Note: [+] $P < .01$. All other loadings are significant, $P < .001$.

TABLE 4: Standardized WLSMV estimates for correlations between theory of mind, age and verbal ability in UK, Japan, and Italy.

	UK	Japan	Italy
Age	.38[*]	.46[*]	.13
Verbal	.66[*]	.65[*]	.47[*]

Note: [*] $P < .001$.

factor exhibited partial measurement invariance [48]. With the exception of one indicator, the theory-of-mind latent factor exhibited equal form, equal factor loadings, and equal thresholds. Table 4 shows the standardized theory-of-mind factor loadings for each nation for the partial measurement invariance model.

3.3. Group Differences in Children's Theory of Mind. Our next aim was to examine group differences in performance on the latent theory-of-mind factor. To ensure that potential group differences in latent factor means were interpretable, we constrained the variance of the latent factor to be equal (0.42) across each of the national groups, $\chi^2(66) = 92.81$, $P = .02$, CFI = .99, TLI = .98, AIC = 6.06. This additional constraint did not result in a significant degradation of model fit, $\Delta\chi^2(2) = 2.88$, $P = .24$. To test overall group differences in the theory-of-mind latent factor, we further constrained the means to be equal across national groups. This constraint decreased the model fit from the unconstrained solution, $\chi^2(68) = 112.53$, $P = .0006$, CFI = .97, TLI = .97, $\Delta\chi^2(2) = 10.85$, $P = .004$, indicating a significant mean contrast between national groups on the latent theory-of-mind factor.

To explore this contrast further, we fixed the mean of the UK theory-of-mind latent factor to zero so that the mean of the other two groups represented deviations from the latent mean of British children. When we applied Bonferroni's adjustment to compensate for multiple comparisons ($\alpha = .02$) we found that children in UK performed significantly better than children in Italy, 0.61 SD, $P = .002$ and marginally better than children in Japan, 0.41 SD, $P = .02$. Next, fixing the mean of the latent factor for Japanese children to zero so that the mean for the Italian children represented the difference, we found no significant mean difference in performance for children in Japan and Italy, 0.20 SD, $P = .29$. In summary,

even when the noninvariance in one item was accounted for, children from the UK outperformed children from the other two countries, with a medium to large effect for the contrast with Italy and a small to medium effect for the contrast with Japan [50].

3.4. Theory of Mind, Age, and Receptive Vocabulary across Groups. Indicators of receptive vocabulary and age were entered into the partial measurement invariance multiple-groups solution and allowed to covary with the theory-of-mind latent factor. This model provided an adequate fit to the data, $\chi^2(108) = 148.18$, $P = .01$, CFI = .98, TLI = .98. From Table 4, which presents the correlations between the theory-of-mind latent factor score and both receptive vocabulary and age by country, it can be seen that there were moderate correlations between age and theory of mind in UK and Japan but not in Italy and strong correlations between receptive vocabulary and theory of mind in UK and Japan but only moderate links between receptive vocabulary and theory of mind in Italy. To test whether the strength of the correlations between the theory-of-mind latent factor, receptive vocabulary, and age was significantly different in each nation, we constrained the correlations between these variables to equality. The model fit degraded significantly against the baseline (unconstrained) solution, $\chi^2(114) = 195.64$, $P = .00001$, CFI = .96, TLI = .95, $\Delta\chi^2(6) = 27.30$, $P = .0001$. Thus, there were significant differences in the strength of the association between individual differences in performance on the theory-of-mind latent factor and age and receptive vocabulary across the three nations.

The baseline (unconstrained) solution revealed significant across group differences in performance on the theory-of-mind latent factor, even when effects of individual differences in age and receptive vocabulary were controlled. Again, we applied Bonferroni's adjustment to compensate for multiple comparisons ($\alpha = .02$). Both Japanese and Italian children performed worse than children from UK: the average difference was 0.43 SD, $P = .02$, for Japanese children and 0.62 SD, $P = .001$ for Italian children. There was no significant difference between Japanese and Italian children, 0.19 SD, $P = .29$. In summary, with effects of age and receptive vocabulary controlled, the group differences remained largely

the same: children from UK outperformed children from the other two countries, with a large effect for the contrast with Italy and a medium effect for the contrast with Japan [51].

4. Discussion

This study compared theory-of-mind task performance in 6-year olds (matched for age, gender, and verbal ability) living in three small but historic cities in UK, Japan, and Italy (Cambridge, Pavia, and Kyoto). Our first aim was to ensure that group comparisons were valid and meaningful; to this end we applied CFAs to establish across-culture measurement invariance. Our second aim was to test two competing hypotheses regarding mean theory-of-mind scores for children from each of these three countries. The first "general culture" hypothesis was that children in the two Western countries (UK and Italy) would outperform children growing up in a collectivistic culture in Japan on tests tapping the awareness of the subjective nature of mental states. The second "pedagogical experience" hypothesis was that children in the UK (who begin school a year earlier than children in Italy or Japan) would outperform the other two groups on the theory-of-mind task battery. Our results supported this second hypothesis, with group differences remaining significant even when effects of verbal ability were controlled.

4.1. Do Theory-of-Mind Tasks Show Measurement Invariance across Cultures? We found that a parsimonious solution in which each theory-of-mind indicator loaded onto a single theory-of-mind factor provided the best fit to our data suggesting that performance on each indicator was underpinned by individual differences in mental-state reasoning. It is possible that each of the theory-of-mind indicators loaded onto this latent factor for theoretically less interesting reasons. For example, the theory-of-mind indicators may have been related to a single latent factor because of shared demands on children's story comprehension or general language skills. Against these more trivial accounts, it is worth noting that while the data supported a single factor solution, the loadings were weaker for the items regarding inference of emotion based on false belief. To establish divergent validity more fully, future research using CFA to measure individual differences in theory of mind would benefit from including items that match the structure of the theory-of-mind tasks but do not involve reasoning about mental states. Such items would not be expected to load significantly on a latent theory-of-mind factor. While the current study did not include such items, we were able to conduct a subsequent analysis with a single factor model for the whole sample in which each of the theory-of-mind indicators loaded onto the single latent factor and was regressed onto receptive vocabulary scores. The resulting model provided a good fit to the data, $\chi^2(18) = 44.99$, $P = .0008$, CFI = .98, TLI = .96, RMSEA = .07. The latent theory-of-mind factor variance was significant, unstandardized estimate = 0.43, $P < .001$, and the mean factor loading was .61, Range: .30–.79, all $P < .01$. Thus,

individual differences in verbal ability could not fully explain the links between these items.

Perhaps more important than assessing the fit of a measurement model, CFA also allows one to minimize spurious results by examining whether a model shows measurement invariance across different groups. Until very recently, cross-cultural psychological studies have paid remarkably little attention to issues of measurement invariance [33]. For example, this study is the first cross-cultural comparison of theory of mind to examine this assumption directly. Our single factor solution showed partial measurement invariance [48], in that 7 of the 8 task indicators exhibited invariance across each of the three cultural groups. Rather than removing this item or ignoring the item's noninvariance, multiple-groups CFA permitted us to release the equality constraints on this item before comparing group means. This approach is less likely than traditional methods (e.g., ANOVA) to introduce bias into estimates of group means [33, 48]. As the partial measurement invariance solution was sufficient to assess group contrasts, we can infer that the majority of theory-of-mind tasks used in our study measured the same construct in the same way in very different cultural settings. In short, the results from our multiple-groups CFA are reassuring in that they indicate that the meaning of these widely used tests is not "lost in translation."

The exceptional item asked children what a character (mistakenly) thought he was getting for his birthday. This item showed differential item functioning (DIF) in that different scores were obtained by children from each of the national groups who actually had the same level of underlying latent theory-of-mind ability [34]. Specifically, while participants in UK found this item easier than would be expected on the basis of their latent theory-of-mind factor scores, participants in the Italian group found this item more difficult than would be expected. Here, it is worth noting that when administering this task to Italian children, one author (SL) observed that children often objected to this story, arguing that "mums do not tell lies to their children" or "mums always tell the truth." These comments suggest that the Italian children may have performed particularly poorly on this task because they found the narrative implausible. To elucidate the origins of this differential item functioning, the relationship between actor and partner in the stories and the motivation for deception, could be varied systematically in future studies in order to establish whether group differences in task performance reflect deontic cultural contrasts in either of these specific elements.

4.2. Explaining Similarities and Contrasts across Cultures. It is remarkable that despite the difference in analytical approach and sample age (our study included 5- to 6-year olds, whereas previous studies have involved younger children), our findings echo those from previous studies [18, 38]. Note also, that in contrast with previous studies, this investigation included comparisons both within the West and between East and West. This feature of the study allowed us to pit two competing hypotheses against each other. In the former "general culture" hypothesis, group differences

were expected between East and West, with children from collectivistic cultures (such as Japan) showing less advanced understanding of the subjective nature of mental states than children from individualistic Western cultures (such as Italy and UK). In the second "pedagogical experiences" hypothesis originally proposed by [38], the advantage in theory-of-mind performance was predicted to be specific to children from UK, who begin school a year before children in Italy or Japan. Consistent with this second hypothesis, we found that the children from UK outperformed both the Japanese and Italian children and that there were no significant differences between the latter two groups.

While the contrast in children's pedagogical experiences provides a simple and plausible account of the group differences obtained in this study, it is important to acknowledge that we did not have access to information about individual children's families, such that several family-focused accounts of cultural difference also deserve consideration. For example, within-culture studies indicate that individual differences in children's false-belief understanding show robust associations with variation in the quality of mother-child relationships [52] and in patterns of mother-child talk [10].

Cross-cultural studies indicate that, in comparison with American (and by inference British) mothers, Italian mothers favor social-oriented interactions [53], adopt a parenting style centered on intimacy, physical affection, and emotional availability [54, 55]. While these contrasts would all appear to favor Italian children, the results from two other studies suggest a different picture. Specifically, Tardif et al. [56] found that Italian mothers talked less often with their toddlers but asked more "test" questions (e.g., "What animal is that?") than British mothers, who asked more genuine questions (e.g., "What would you like to do?"). This contrast in the proportion of genuine questions is interesting and indicates that British mothers show a greater tendency to consider their children as thinking individuals (i.e., as autonomous agents). Consistent with this view, recent cross-cultural work on the social origins of childhood anxiety has demonstrated that Italian mothers are more intrusive and controlling and less autonomy granting than British mothers [57].

Finally, while we did not have access to information about family size for our study participants, it is worth noting that the three countries in this study also differ in mean number of children per family. Recent data released by the OECD [58] indicates that, among children aged 0–14 years, the percentage of children with one or more sibling is 78% in UK, 75% in Italy, and 72% in Japan (with corresponding fertility rates of 1.94, 1.41 and 1.37). In Western samples, the presence of siblings, especially older siblings, has been found to predict false-belief performance (e.g., [59]). Interestingly, however, this facilitative effect of siblings is not evident among Japanese children [12], and for Chinese children (who typically do not have siblings), contact with cousins (i.e., with other children in the family) appears negatively related to false-belief performance [60]. This contrast may reflect cultural differences in the *nature* of child-child relationships, as older children in collectivist cultures are strongly encouraged to take on a caregiving role and so may not be as playful in their interactions as children in more individualistic cultures

[61]. Such qualitative contrasts in the nature of the sibling relationship in different cultures are likely to be important, as other studies have shown that, for preschoolers, the advantage of having an older sibling is restricted to siblings aged under 12 who, one might presume, are more likely to act as playmates for preschooler than are older children [59, 62].

At this point it is worth noting that although East-West contrasts in executive function have received considerable attention from researchers (e.g., [22]), a recent meta-analysis of data from 10,000 children from 15 different countries has shown no geographical contrasts in the association between false-belief performance and executive function [63]. Moreover, as noted by Liu et al. [18], a myriad of factors are likely to contribute to cultural contrasts in performance (e.g., exposure to formal schooling, quality, and content of family interactions). Future research is needed to test more complex accounts of between-country contrasts. For example, exposure to formal schooling with the associated requirements for children to regulate their behaviour and attend to instructions may accelerate the maturation of executive functions and so indirectly enhance children's growing awareness of mental states.

4.3. Conclusions and Caveats. We propose that the findings from this study contribute to the literature both methodologically and conceptually. At a methodological level, the careful matching of samples for verbal ability (as well as age and gender) made this study more rigorous than many previous studies of cultural differences in children's acquisition of a concept of mind. In addition, by adopting a latent variable approach we were able to conduct multigroup CFAs in order to test directly whether spurious measurement effects were likely to have contributed to cultural contrasts reported in previous studies. In each of these respects, our findings were reassuring. Even though the groups were more carefully matched, our results were strikingly similar to those reported in previous studies; moreover, the latent factors obtained in each country showed a similar structure, with similar factor loadings and similar factor variance.

Demonstrating the similarities in factor structure, loading, and variance also has conceptual implications. Specifically, our results indicate that between-country contrasts reported in previous studies are unlikely to be a spurious artefact of measurement effects as, of the 8 items used, just one was noninvariant and this item had not been used in previous comparisons of samples from the East and West. That said, although latent mean comparisons can be carried out even when some items are noninvariant [48], our findings can only be viewed as a first step towards elucidating cultural differences in children's theory of mind. In particular, although our finding of measurement equivalence addresses concerns about the validity of applying measures of theory of mind in different languages with different cultures, there remains much work to be done in addressing other concerns about cross-cultural comparisons (e.g., whether variation in task performance has the same origins and consequences for children from different countries).

Another key conceptual finding was that the East-West contrast was restricted to children in UK, with the largest group difference being observed between children in UK and children in Italy. Thus, our findings challenge the "general culture" hypothesis and suggest that specific experiences (perhaps particularly children's experiences of formal schooling) may have greater impact on children's developing understanding of mental states. However, a key limitation of the current study was the lack of direct information about children's conversations and relationships both within and outside the family. Two important goals for future research in this field are (i) recruiting samples that enable more refined comparisons (e.g., from cultures that speak the same language but differ in the age at which children begin formal schooling) and (ii) direct assessing aspects of children's social environments (both at home and at school) in order to locate more precisely the factors that contribute to between-country differences in the rate at which children acquire an understanding of their own and others' minds.

Conflict of Interests

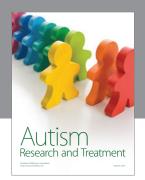The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] C. Blair and R. P. Razza, "Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten," *Child Development*, vol. 78, no. 2, pp. 647–663, 2007.

[2] S. Lecce, M. Caputi, and C. Hughes, "Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement," *Journal of Experimental Child Psychology*, vol. 110, no. 3, pp. 313–331, 2011.

[3] C. Hughes, R. Ensor, and A. Marks, "Individual differences in false belief understanding are stable from 3 to 6 years of age and predict children's mental state talk with school friends," *Journal of Experimental Child Psychology*, vol. 108, no. 1, pp. 96–112, 2011.

[4] R. Banerjee, D. Watling, and M. Caputi, "Peer relations and the understanding of faux pas: longitudinal evidence for bidirectional associations," *Child Development*, vol. 82, no. 6, pp. 1887–1905, 2011.

[5] M. Caputi, S. Lecce, A. Pagnin, and R. Banerjee, "Longitudinal effects of theory of mind on later peer relations: the role of prosocial behaviour," *Developmental Psychology*, vol. 48, pp. 257–270, 2012.

[6] C. Hughes, *Social Understanding, Social Lives: From Toddlerhood Through To the Transition to School*, Psychology Press, London, UK, 2011.

[7] G. Pavarini, D. de Hollanda Souza, and C. Hawk, "Parental practices and theory of mind development," *Journal of Child and Family Studies*, vol. 22, pp. 844–853.

[8] J. Perner, T. Ruffman, and S. Leekam, "Theory of mind is contagious: you catch it from your sibs," *Child Development*, vol. 65, pp. 1228–1238, 1994.

[9] C. Lewis, N. H. Freeman, C. Kyriakidou, K. Maridaki-Kassotaki, and D. M. Berridge, "Social influences on false belief access: specific sibling influences or general apprenticeship?" *Child Development*, vol. 67, no. 6, pp. 2930–2947, 1996.

[10] R. Ensor and C. Hughes, "Content or connectedness? Mother-child talk and early social understanding," *Child Development*, vol. 79, no. 1, pp. 201–216, 2008.

[11] R. Ensor, R. T. Devine, A. Marks, and C. Hughes, "Mothers' cognitive references to two-year-olds predict theory of mind at ages 6 and 10," *Child Development*, 2013.

[12] T. Ruffman, L. Slade, and E. Crowe, "The relation between children's and mothers' mental state language and theory-of-mind understanding," *Child Development*, vol. 73, no. 3, pp. 734–751, 2002.

[13] J. Dunn, "Children as psychologists: the later correlates of individual differences in understanding of emotions and other minds," *Cognition and Emotion*, vol. 9, pp. 187–201, 1995.

[14] C. Hughes, K. K. Fujisawa, R. Ensor, S. Lecce, and R. Marfleet, "Cooperation and conversations about the mind: a study of individual differences in 2-year-olds and their siblings," *British Journal of Developmental Psychology*, vol. 24, no. 1, pp. 53–72, 2006.

[15] C. Hughes and J. Dunn, "Understanding mind and emotion: longitudinal associations with mental-state talk between young friends," *Developmental Psychology*, vol. 34, no. 5, pp. 1026–1037, 1998.

[16] J. Avis and P. Harris, "Belief-desire reasoning among Baka children: evidence for a universal conception of mind," *Child Development*, vol. 62, pp. 460–467, 1991.

[17] T. Callaghan, P. Rochat, A. Lillard et al., "Synchrony in the onset of mental-state reasoning: evidence from five cultures," *Psychological Science*, vol. 16, no. 5, pp. 378–384, 2005.

[18] D. Liu, H. M. Wellman, T. Tardif, and M. A. Sabbagh, "Theory of mind development in Chinese children: a meta-analysis of false-belief understanding across cultures and languages," *Developmental Psychology*, vol. 44, no. 2, pp. 523–531, 2008.

[19] F. Shinagawa, S. Kobayashi, K. Fujita, and H. Maekawa, *Nihon-ban WAIS-R Seijin Chinou Kensahou*, Nihon Bunka Kagakusha, Tokyo, Japan, 1990.

[20] R. T. Devine and C. Hughes, "Relations between false-belief understanding and executive function in early childhood: a meta-analysis," *Child Development*. In press.

[21] A. Shahaeian, C. C. Peterson, V. Slaughter, and H. M. Wellman, "Culture and the sequence of steps in theory of mind development," *Developmental Psychology*, vol. 47, no. 5, pp. 1239–1247, 2011.

[22] C. Lewis, M. Koyasu, S. Oh, A. Ogawa, B. Short, and Z. Huang, "Culture, executive function, and social understanding," *New Directions for Child and Adolescent Development*, vol. 2009, no. 123, pp. 69–85, 2009.

[23] S. Oh and C. Lewis, "Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding," *Child Development*, vol. 79, no. 1, pp. 80–99, 2008.

[24] M. Koyasu, "Can visual feedback effect perspective-taking behavior in young children?" *Psychologia*, vol. 40, no. 2, pp. 91–103, 1997.

[25] M. Naito and K. Koyama, "The development of false-belief understanding in Japanese children: delay and difference?" *International Journal of Behavioral Development*, vol. 30, no. 4, pp. 290–304, 2006.

[26] A. Lillard, "Ethnopsychologies: cultural variations in theories of mind," *Psychological Bulletin*, vol. 123, no. 1, pp. 3–32, 1997.

[27] H. M. Wellman, F. Fang, and C. C. Peterson, "Sequential Progressions in a theory-of-mind scale: longitudinal perspectives," *Child Development*, vol. 82, no. 3, pp. 780–792, 2011.
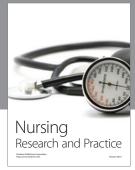
[28] C. Hughes and J. Dunn, "Understanding mind and emotion: longitudinal associations with mental-state talk between young friends," *Developmental Psychology*, vol. 34, no. 5, pp. 1026–1037, 1998.

[29] P. Harris, C. Johnson, D. Hutton, G. Andrews, and T. Cooke, "Young children's theory of mind and emotion," *Cognition and Emotion*, vol. 3, pp. 379–400, 1989.

[30] S. A. Miller, "Children's understanding of second-order mental states," *Psychological Bulletin*, vol. 135, no. 5, pp. 749–773, 2009.

[31] K. Milligan, J. W. Astington, and L. A. Dack, "Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding," *Child Development*, vol. 78, no. 2, pp. 622–646, 2007.

[32] B. M. Byrne and T. L. Campbell, "Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: a look beneath the surface," *Journal of Cross-Cultural Psychology*, vol. 30, no. 5, pp. 555–574, 1999.

[33] S. Lecce, M. Caputi, and C. Hughes, "Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement," *Journal of Experimental Child Psychology*, vol. 110, no. 3, pp. 313–331, 2011.

[34] T. Brown, *Confirmatory Factor Analysis for Applied Research*, The Guilford Press, New York, NY, USA, 1st edition, 2006.

[35] Cambridgeshire Country Council, "Household Income," Cambridgeshire, 2009, http://www.cambridgeshire.gov.uk/business/research/economylab/labour/houseincome.htm.

[36] Ministry for Internal Affairs and Communications, "Chapter 13: Family budgets and prices," 2013, http://www.stat.go.jp/english/data/handbook/c0117.htm#c13.

[37] G. di Ghistanzoni Cardoni, "Camera di Commercio Pavia," 2013, http://www.pv.camcom.it/index.phtml?Id_VMenu=1.

[38] R. Ensor, R. Devine, and C. Hughes, "Mothers' cognitive references to two-year-olds predicts theory of mind at ages 6 and 10," *Child Development*.

[39] C. Hughes, A. Adlam, F. Happé, J. Jackson, A. Taylor, and A. Caspi, "Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 41, no. 4, pp. 483–490, 2000.

[40] C. Hughes and S. Lecce, "The Italian Job? Comparing theory of mind in British and Italian children," *British Journal of Developmental Psychology*, vol. 28, pp. 747–7766, 2010.

[41] J. Perner and H. Wimmer, ""John thinks that Mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children," *Journal of Experimental Child Psychology*, vol. 39, no. 3, pp. 437–471, 1985.

[42] K. Sullivan, D. Zaitchik, and H. Tager-Flusberg, "Preschoolers can attribute second-order beliefs," *Developmental Psychology*, vol. 30, pp. 395–402, 1994.

[43] E. Meins, C. Fernyhough, R. Wainwright, M. Das Gupta, E. Fradley, and M. Tuckey, "Maternal mind-mindedness and attachment security as predictors of theory of mind understanding," *Child Development*, vol. 73, no. 6, pp. 1715–1726, 2002.

[44] T. Tardif, M. Shatz, and L. Naigles, "Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin," *Journal of Child Language*, vol. 24, no. 3, pp. 535–565, 1997.

[45] D. R. Divgi, "Calculation of the tetrachoric correlation coefficient," *Psychometrika*, vol. 44, no. 2, pp. 169–172, 1979.

[46] D. Wechsler, *The Wechsler Preschool and Primary Scale of Intelligence*, The Psychological Corporation, London, UK, 3rd edition, 2003.

[47] R. Kline, *Principles and Practice of Structural Equation Modeling (Methodology in the Social Sciences)*, Guilford Press, New York, NY, USA, 3rd edition, 2012.

[48] B. M. Byrne, R. J. Shavelson, and B. Muthén, "Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance," *Psychological Bulletin*, vol. 105, no. 3, pp. 456–466, 1989.

[49] L. K. Muthen and B. O. Muthen, *MPlus: Statistical Analysis With Latent Variables User Guide*, Muthen and Muthen, Los Angeles, Calif, USA, 6th edition, 2010.

[50] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Erlbaum, Hillsdale, NJ, USA, 1988.

[51] J. Cohen, "A power primer," *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992.

[52] E. Meins, C. Fernyhough, R. Wainwright, M. Das Gupta, E. Fradley, and M. Tuckey, "Maternal mind-mindedness and attachment security as predictors of theory of mind understanding," *Child Development*, vol. 73, no. 6, pp. 1715–1726, 2002.

[53] M. H. Bornstein, L. R. Cote, and P. Venuti, "Parenting beliefs and behaviors in northern and southern groups of Italian mothers of young infants," *Journal of Family Psychology*, vol. 15, no. 4, pp. 663–675, 2001.

[54] M. H. Bornstein, D. L. Putnick, J. T. D. Suwalsky et al., "Emotional relationships in mothers and infants: culture-common and community-specific characteristics of dyads from rural and metropolitan settings in Argentina, Italy, and the United States," *Journal of Cross-Cultural Psychology*, vol. 43, no. 2, pp. 171–197, 2012.

[55] H.-C. Hsu and M. Lavelli, "Perceived and observed parenting behavior in American and Italian first-time mothers across the first 3 months," *Infant Behavior and Development*, vol. 28, no. 4, pp. 503–518, 2005.

[56] J. Tobin, Y. Hsueh, and M. Karasawa, *Preschool in Three Cultures Revisited: China, Japan, and the United States*, The University of Chicago Press, Chicago, Ill, USA, 2009.

[57] T. Ruffman, J. Perner, M. Naito, L. Parkin, and W. A. Clements, "Older (but not younger) siblings facilitate false belief understanding," *Developmental Psychology*, vol. 34, no. 1, pp. 161–174, 1998.

[58] OECD, "Families and Children: Family Database," 2012, http://www.oecd.org/els/family/oecdfamilydatabase.htm.

[59] T. Ruffman, J. Perner, M. Naito, L. Parkin, and W. A. Clements, "Older (but not younger) siblings facilitate false belief understanding," *Developmental Psychology*, vol. 34, no. 1, pp. 161–174, 1998.

[60] C. Lewis, Z. Huang, and M. Rooksby, "Chinese preschoolers' false belief understanding: is social knowledge underpinned by parental styles, social interactions or executive functions?" *Psychologia*, vol. 49, no. 4, pp. 252–266, 2006.

[61] L. Wei, *Scrap Live-in Rule for Maids, Say Advocacy Groups*, South China Morning Post, 2013.

[62] A. McAlister and C. Peterson, "A longitudinal study of child siblings and theory of mind development," *Cognitive Development*, vol. 22, no. 2, pp. 258–270, 2007.

[63] R. T. Devine and C. Hughes, "Relations between false-belief understanding and executive function in early childhood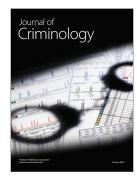: a meta-analysis," *Child Development*. In press.