# Theory of Mind Development in Chinese Children: A Meta-Analysis of False-Belief Understanding Across Cultures and Languages

4 authors, including:

David Liu
HuaiHai Institute of Technology
19 PUBLICATIONS   2,659 CITATIONS

Twila Tardif
University of Michigan
99 PUBLICATIONS   3,826 CITATIONS

Mark A Sabbagh
Queen's University
79 PUBLICATIONS   3,813 CITATIONS

# Theory of Mind Development in Chinese Children: A Meta-Analysis of False-Belief Understanding Across Cultures and Languages

David Liu, Henry M. Wellman, and Twila Tardif
University of Michigan

Mark A. Sabbagh
Queen's University at Kingston

Theory of mind is claimed to develop universally among humans across cultures with vastly different folk psychologies. However, in the attempt to test and confirm a claim of universality, individual studies have been limited by small sample sizes, sample specificities, and an overwhelming focus on Anglo-European children. The current meta-analysis of children's false-belief performance provides the most comprehensive examination to date of theory-of-mind development in a population of non-Western children speaking non-Indo-European languages (i.e., Mandarin and Cantonese). The meta-analysis consisted of 196 Chinese conditions (127 from mainland China and 69 from Hong Kong), representing responses from more than 3,000 children, compared with 155 similar North American conditions (83 conditions from the United States and 72 conditions from Canada). The findings show parallel developmental trajectories of false-belief understanding for children in China and North America coupled with significant differences in the timing of development across communities—children's false-belief performance varied across different locales by as much as 2 or more years. These data support the importance of both universal trajectories and specific experiential factors in the development of theory of mind.

Keywords: theory of mind, false belief, meta-analysis, Chinese, culture

Everyday attribution of mental states—a "theory of mind"—distinguishes human social cognition from that of other animals, including nonhuman primates. Human theory of mind is not only phylogenetically distinctive, it is arguably ontogenetically universal. Except in cases of neurodevelopmental disorders, such as autism (Baron-Cohen, 1995), or in cases of severely restricted linguistic inputs, such as deaf children of nonsigning parents (Peterson & Siegal, 2000), the claim is that everyone develops a theory of mind (Scholl & Leslie, 2001; Wellman, 1998). Confirming and characterizing universality in theory of mind, however, requires the collection of cross-cultural data, which poses several challenges with regard to appropriate measures and samples.

The most widely used measures of theory-of-mind development are false-belief tasks. One version involves an unseen change of location (Wimmer & Perner, 1983): A child sees Maxi put his chocolate in a kitchen cupboard and then leave. While Maxi is outside, his mother moves the chocolate from the cupboard to a drawer. Maxi returns, and the child is asked, "Where does Maxi *think* the chocolate is?" or "Where will Maxi *look* for his chocolate?" Children develop from answering these sorts of questions according to reality to answering according to *false belief*. This developmental pattern is consistent across numerous task modifications, thereby establishing that false-belief tasks measure a critical conceptual development (Wellman, Cross, & Watson, 2001).

Whether false-belief understanding shows a similar trajectory and/or timetable across disparate cultures has not been thoroughly and convincingly addressed to date because the bulk of the studies have been with children in communities dominated by Western European cultures, speaking Indo-European languages. Recently, a large number of studies (both published and unpublished) have investigated the emergence of false-belief understanding in Chinese children. Indeed, false-belief research in China now constitutes the largest sample of non-Western data. This large body of work provides a unique opportunity to use powerful meta-analytic procedures to evaluate whether the character of development in false-belief understanding is similar across divergent cultural and linguistic communities.

Although many researchers have argued that Western cultures' conception of mind is not always wholly shared by other cultures (Lillard, 1998; Shweder & Bourne, 1984), only a few studies have directly compared false-belief performance between Western and non-Western communities (e.g., Callaghan et al., 2005; Vinden, 1999; Wellman et al., 2001). These studies came to inconsistent conclusions about developmental trajectory and timing. Callaghan et al. (2005) tested a modest number of children (from 13 to 31 children per age for each locale) on a single task across five

countries (Canada, India, Peru, Samoa, and Thailand). They found a universal trajectory from below- to above-chance performance and, although they did not conduct any statistical analyses directly comparing data between locales, concluded that there is a tightly synchronous timetable for the onset of false-belief competence. Wellman et al. (2001) conducted a meta-analysis comparing false-belief performance across age, various tasks, and countries. They argued that the trajectory from below- to above-chance performance is universal but (in contrast to Callaghan et al.) also argued that the timing for this development varies across countries. These prior findings are intriguing, but significant questions remain. In particular, all prior analyses involved small samples from outside Western European cultures, which did not allow for precise estimation of developmental timetables across cultures. The current meta-analysis encompasses a large number of non-Western children to better compare the acquisition timetables of false-belief understanding.

Chinese children provide an excellent non-Western comparison because folk psychologies, societal expectations, and parental practices differ significantly between Chinese and Western cultures (Nisbett, Peng, Choi, & Norenzayan, 2001; Wang, 2004), all of which could potentially affect children's developing understanding of mental states. Moreover, there are cultural differences that are directly relevant to factors that are known to influence theory-of-mind development. For one, exposure to mental state verbs, such as *think*, *know*, and *want*, influences theory-of-mind development among Western children (Bartsch & Wellman, 1995; Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991; Ruffman, Slade, & Crowe, 2002). However, verb meanings and acquisition of verbs differ for Chinese versus English (Tardif, 2006). Whereas English terms for talking about belief, such as *think* and *believe*, are neutral with respect to the belief's truthfulness (Moore, Bryant, & Furrow, 1989), Chinese languages contain belief terms that specifically mark someone's belief as false. For Mandarin-speaking adults, the term *yi3wei2* denotes false belief (and the term *dang1* can also have connotations of false belief), whereas *xiang3* is neutral (Lee, Olson, & Torrance, 1999). Similarly, in Cantonese, the term *ji5wai4* denotes false belief, whereas *soeng2* and *nam5* are neutral (Tardif, Wellman, & Cheung, 2004). Second, research has shown that within Western samples, executive functioning correlates with performance on false-belief tasks (Carlson & Moses, 2001). Impulse control is especially valued by Chinese parents and appears early in Chinese children (X. Chen et al., 1998), and Chinese children on the mainland (Sabbagh, Xu, Carlson, Moses, & Lee, 2006) and in Hong Kong (Tardif, So, & Kaciroti, 2007) outperform American children in executive functioning.

For our meta-analysis, we assembled published and unpublished studies encompassing almost 200 conditions (typically 16–24 children per condition) from mainland China (Mandarin speaking) and Hong Kong (Cantonese speaking). For comparison, we included available studies from North America (the United States and Canada), which constitutes the largest sample of Western data available. Note that we are not specifically predicting earlier or different development for one group or another. More simply, we argue for the value of comprehensive data from non-Western children and outline factors that make Chinese children an important comparison case.

## Method

### Unit of Analysis

Conditions within studies, rather than individual participants or entire studies, make up the unit of analysis in a meta-analysis (Glass, McGraw, & Smith, 1981). Our dependent variable was the proportion of false-belief questions answered correctly in each condition. A meta-analysis of such data (readily available in false-belief studies) circumvents the use of effect sizes as the dependent variable, circumventing problems with the appropriateness, assumptions, and robustness of various techniques of pooling different indirect measures of effect size.

### Studies

We first searched relevant databases, journals, and abstracts of recent conferences for studies on theory of mind, false belief, belief reasoning, understanding mental states, and folk or naïve psychology in China. Because a potential weakness of any meta-analysis is publication bias against certain, usually nonsignificant, results, we contacted Chinese researchers with as yet unpublished data. We ended our search for relevant studies in April 2004.

Of the studies found, we included only those with false-belief conditions with typically developing Chinese children (not autistic or other delayed samples). We did not include conditions with Chinese children outside of China (e.g., in Taiwan or in Europe). For the North American data, we used the conditions from Wellman et al.'s (2001) meta-analysis that were from the United States and Canada. We selected only conditions that matched the Chinese data in task characteristics. For example, none of the Chinese conditions involved deception, so we did not include any North American conditions that involved deception. As shown in Table 1, we included 196 separate Chinese conditions (127 from mainland China and 69 from Hong Kong) and 155 separate North American conditions (83 from the United States and 72 from Canada).

### Coding

Each condition was coded for the proportion of false-belief judgments answered correctly and for various features making up the independent variables. Some variables used by Wellman et al. (2001) were not included because they were unavailable in the Chinese data (e.g., deception); some levels of a few variables were truncated because there were no or few conditions with those levels. The resulting coding scheme was

1. Age: Mean age (in months) of children in a condition.

2. Locale of participants: Mainland China, Hong Kong, the United States, or Canada.

3. Type of false-belief task: Change of location, unexpected contents, or deceptive identity (use of appearance–reality stimuli in a false-belief task).

4. Nature of the protagonist: A puppet or doll, a pictured character, or a real person (present or absent). Protagonists as real persons were also coded as either the self or another person.

Table 1

*Listing of the Studies and Conditions in the Meta-Analysis*

| Location and study | Year | Total conditions included in meta-analysis | Lab group |
|---|---|---|---|
| Mainland China | | | |
| M. J. Chen & Lin | 1994 | 4 | Other |
| Deng | 2001 | 32 | Deng |
| Goetz | 2003 | 8 | Other |
| Lee et al. | 1999 | 36 | Lee |
| Sabbagh et al. | 2006 | 9 | Other |
| Sang et al. | 2004 | 12 | Deng |
| Tardif et al. | 2000 | 12 | Tardif |
| Tardif et al. | 2001 | 12 | Tardif |
| Volling et al. | 1999 | 2 | Other |
| Hong Kong | | | |
| Tardif et al. | 2001 | 12 | Tardif |
| Tardif et al. | 2004 | 21 | Tardif |
| Tardif et al. | 2007 | 24 | Tardif |
| Tardif & Ng | 2001 | 12 | Tardif |
| Total conditions from China | | 196 | |
| United States | | | |
| Bartsch & Wellman | 1989 | 1 | |
| Carlson et al. | 1998 | 2 | |
| Dalke | 1995 | 14 | |
| Davis | 1997 | 3 | |
| Frye et al. | 1995 | 12 | |
| Hickling et al. | 1997 | 2 | |
| Lillard & Flavell | 1992 | 3 | |
| Moses | 1993 | 2 | |
| Robinson & Mitchell | 1995 | 8 | |
| Sheffield et al. | 1993 | 2 | |
| Slaughter & Gopnik | 1996 | 4 | |
| Sullivan & Winner | 1991 | 12 | |
| Sullivan & Winner | 1993 | 10 | |
| Taylor & Carlson | 1997 | 4 | |
| Winner & Sullivan | 1993 | 2 | |
| Zaitchik | 1990 | 2 | |
| Canada | | | |
| Astington et al. | 1989 | 8 | |
| Carpendale & Chandler | 1996 | 13 | |
| Chandler & Hala | 1994 | 2 | |
| Gopnik & Astington | 1988 | 42 | |
| Hala et al. | 1991 | 3 | |
| Moore et al. | 1990 | 2 | |
| Ruffman et al. | 1993 | 2 | |
| Total conditions from North America | | 155 | |

*Note.* The Chinese data came from numerous studies but were often conducted by a handful of collaborative lab groups. Lab groups that contributed 20 or more conditions are identified by the name of the primary collaborator.

5. Nature of the target object: A real or representative object or a pictured object.

6. Salience of the protagonist's mental state: The mental state had to be inferred from the protagonist's simple absence during the key transformation, or the mental state was demonstrated initially on the children themselves (e.g., children initially discovered that the candy box contained pencils).

7. Temporal marker: The false-belief question explicitly marked the time frame (e.g., "When Maxi returns for his chocolate, where will he first look for his chocolate?"), or did not.

8. Type of question: The false-belief question asked children to judge what the protagonist will think or know (excluding conditions using think-falsely verbs) or to judge how the protagonist will behave (e.g., "Where will Maxi look?").

9. Chinese verb form: The false-belief question used a think-falsely verb or a more neutral verb.

## Results

Here we present the results from analyzing the full set of assembled data, which includes both published and unpublished data. However, to preview, all of the effects found to be significant for the full dataset were also found to be significant when only the published data were analyzed (93 of the 127 conditions from mainland China, 45 of the 69 conditions from Hong Kong, 76 of the 83 conditions from the United States, and 62 of the 72 conditions from Canada came from studies that have been published in peer-reviewed journals). Likewise, all of the effects found to be nonsignificant were replicated in both analyses.

### Chinese Studies

We first considered the Chinese studies to characterize the regularities in that data necessary for careful comparisons across locales and languages. We began with the 152 Chinese conditions that asked children to judge someone else's false belief. As shown in Figure 1A, Chinese children's false-belief performance improves dramatically with age. Because all the conditions used tasks with two possible responses, chance equals 50%. Figure 1B shows the same data with the dependent variable, proportion correct, transformed via a logit transformation to allow for examination of the data via linear regression. Chinese children's answers develop from being incorrect (27% correct around the age of 39 months) to being increasingly correct.

The Chinese data came from numerous studies but were often conducted by a handful of collaborative lab groups. Lab groups that contributed 20 or more conditions are identified by the name of the primary collaborator in Table 1. We included lab group as a blocking factor (covariate) entered first in our regression analyses to minimize the influence of data from any single lab. The lab group factor was significant with the 152 Chinese conditions that asked children to judge someone else's false belief, $F(3, 148) = 18.24, p < .001$. However, all of the effects found to be significant with lab group as a blocking factor were also found to be significant without the blocking factor, except for temporal marker.

The first row of Table 2 summarizes the effect of age alone on performance (blocked by lab group); age accounted for 36% of the variance in children's performance. The far right column reports the measure of effect size, computed as an odds ratio. The odds of being correct increase 4.32 times for every 1-year increase in age.

We next examined the effects of the independent variables on this age trend. First consider possible patterns of results: (a) Levels of an additional variable could have no effect on developmental trajectory or timing; (b) an additional variable could be significant
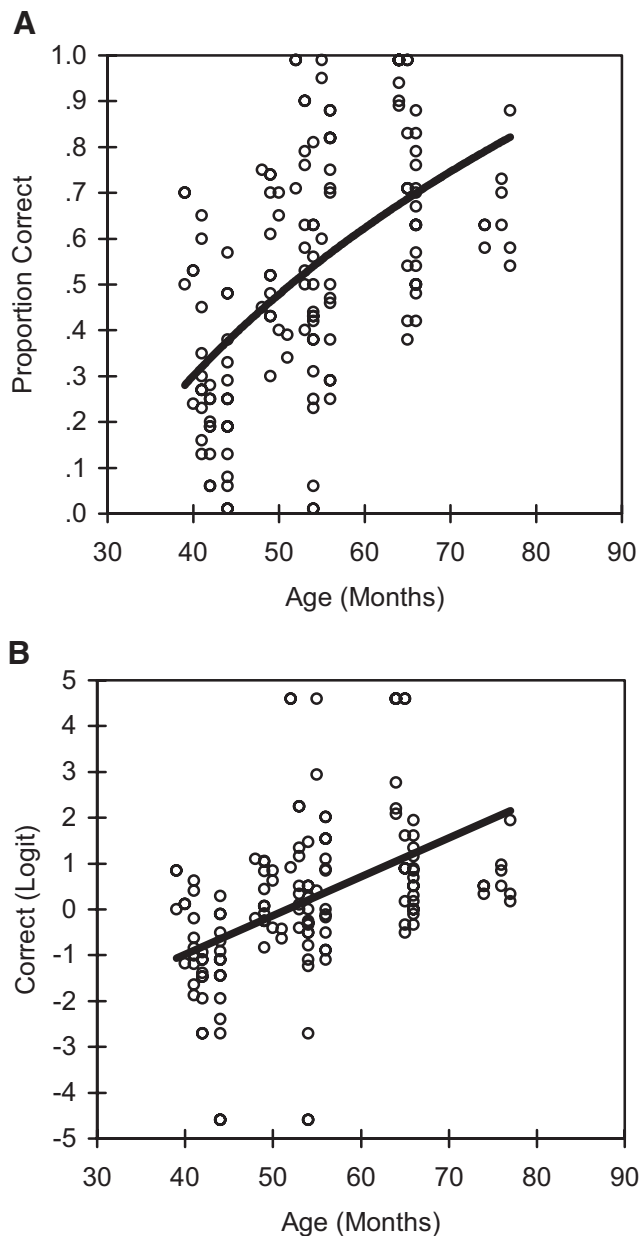
**A**



**B**



*Figure 1.* Scatterplots of Chinese conditions with increasing age showing best-fit line: (A) raw scatterplot of mean proportion correct data with log fit; (B) scatterplot of logit transformed data with linear fit.

only as a main effect without an interaction with age—the timing differs, but the trajectories remain equivalent; (c) the variable could interact significantly with age—levels of that variable change not simply the timing but also the trajectory of development. To assess these possibilities, we used linear regression to check for all two-way interactions with age. For variables that did not interact with age, the interaction term was dropped from the regression and main effects were tested.

As summarized in Table 2, none of the variables we tested interacted significantly with age. In addition, six of the eight task variables had no significant influence on performance. For exam-

ple, performance is the same for any given age whether the protagonist is presented as a real person, a puppet or doll, or a pictured storybook character, and performance is the same across the three types of false-belief tasks: change-of-location tasks, unexpected-contents tasks, and deceptive-identity tasks. To examine possible self versus other differences, we compared conditions that asked children about their own beliefs to conditions that asked them to judge others' beliefs. All self conditions in China used unexpected-contents or deceptive-identity tasks. Therefore, we included only conditions with those tasks in comparing judgments for self (44 conditions) versus other (64 conditions). As reported in Table 2, Chinese children's false-belief judgments of themselves and of others follow identical developmental trajectories. For all these nonsignificant variables, the Chinese data replicate those from Western results (Wellman et al., 2001).

Two task variables, temporal marker and Chinese verb form, were significant as main effects, as listed in Table 2. Use of an explicit temporal marker hindered children's performance. Because this is counter to the effect typically found when such modifications are provided in English (Wellman et al., 2001), it serves as a reminder that different languages work differently; a change that helps clarify linguistic intent in one language may add difficulty in another language. Nevertheless, as discussed above, temporal marker was the only significant effect that was not significant without lab group as a blocking factor, and as we show in the combined regression model analyses below, the temporal marker effect is not robust.

More important, Chinese verb form yielded a main effect. Chinese verb form captures whether the false-belief question used a think-falsely verb or a more neutral verb. This finding indicates that the use of think-falsely verbs in the experimental protocol enhances performance for children of all ages. As demonstrated with the absence of an interaction, children proceed from below- to above-chance performance with increasing age, even with think-falsely verbs in the false-belief tasks.

### Mainland China Versus Hong Kong

Chinese conditions in our dataset encompass two different Chinese languages—Cantonese and Mandarin—and they come from different geographical locales—mainland China (primarily from Beijing but also from Shanghai, Hangzhou, and Wenzhou) and Hong Kong. As shown in Table 3, there was a significant main effect of mainland China versus Hong Kong but no interaction of locale with age. Thus, the trajectory of development is the same in both locales, with timing earlier for mainland Chinese children. Because only one lab group (Tardif's) provided conditions from Hong Kong, we also compared mainland China with Hong Kong using only conditions from this particular lab group. A main effect of mainland China versus Hong Kong, $F(1, 66) = 5.91, p < .05$, remains even with this more controlled analysis.

We constructed a combined regression model with the significant variables (age, Chinese verb form, temporal marker, and mainland China versus Hong Kong) to best predict false-belief performance and to address the possibility that a variable's significant effect may disappear if other significant variables are controlled for as well. In this first multivariate combined model ($R^2 = .679$), temporal marker was no longer significant. Therefore, in the final combined model, we excluded temporal marker; together,

Table 2
*Summary of Results From Meta-Analysis With Chinese Data*

| Variable | Main effect | Interaction with age | Effect size[a] |
|---|---|---|---|
| Age: $F(1, 147) = 145.55, p < .001$ | | | 4.32 for 1 year |
| Nonsignificant | | | |
| Nature of the protagonist | $F(2, 145) = 1.07, p > .30$ | $F(2, 143) = 1.35, p > .20$ | |
| Nature of the target object | $F(1,146) = .50, p > .40$ | $F(1, 145) = 2.75, p > .10$ | |
| Salience | $F(1, 146) = .59, p > .40$ | $F(1, 145) = .02, p > .50$ | |
| Type of question | $F(1, 58) = .80, p > .30$ | $F(1, 57) = .42, p > .50$ | |
| Type of task | $F(2, 145) = 1.26, p > .20$ | $F(2, 143) = 1.51, p > .20$ | |
| Self vs. other | $F(1, 102) = .95, p > .30$ | $F(1, 101) = .28, p > .50$ | |
| Main effects | | | |
| Temporal marker | $F(1, 146) = 5.85, p < .05$ | $F(1, 145) = .26, p > .50$ | 1.76 |
| Chinese verb form | $F(1, 146) = 13.20, p < .001$ | $F(1, 145) = 1.68, p > .10$ | 1.97 |

[a] Effect sizes are presented only for significant variables; effect sizes were computed as odds ratios.

age, Chinese verb form, and mainland China versus Hong Kong accounted for 67.5% of the variance in Chinese children's false-belief performance.

### Cross-Locale Comparisons

The regularities found in the Chinese data allow us to directly and comprehensively compare performance across cultures and languages. Given a main effect of locale within the Chinese conditions, we compared Chinese and North American conditions in terms of four locales: mainland China, Hong Kong, the United States, and Canada. Just as with the Chinese data, the vast majority of North American conditions (61 from the US and 45 from Canada) asked children to judge someone else's false belief. As listed in Table 3, there was a significant main effect of the four locales and no interaction with age. Looking at the effect sizes and Figure 2, it is clear that the developmental trajectories are all very similar, because locale does not interact with age. However, there was a significant pattern of timing differences across the four locales. Specifically, the timing of development for children in mainland China and children in the United States is very similar (an odds ratio effect size close to 1.0 translates into even odds of mainland Chinese and U.S. children being correct at any age). However, Canadian children develop earlier than mainland Chinese and U.S. children, whereas Hong Kong Chinese children develop much later. The differences in timing are clearly shown in

Figure 2. Whereas Canadian children start performing above chance around 38 months of age, Hong Kong children do so around 64 months—a difference of more than 2 years.

We demonstrated earlier that Chinese children perform better on conditions with think-falsely verbs; would locales still show differences in timing if only Chinese think-falsely conditions were compared with the North American conditions? To address this, we compared Chinese conditions with think-falsely verbs (58 from mainland China and 30 from Hong Kong) with conditions from the United States and Canada. As listed in Table 3, even when only Chinese think-falsely conditions—the "best" Chinese data—were included, the main effect of locale remained significant, and there was no significant interaction with age. Furthermore, the effect sizes remained very similar.

### Discussion

Our meta-analysis provides the most comprehensive test to date of theory-of-mind development in non-Western cultures with children speaking non-Indo-European languages. The findings add substantially to claims about the universal, early development of theory of mind. Specifically, we demonstrate parallel developmental trajectories but substantially different timetables across numerous locales. Because we included numerous unpublished as well as published studies, our findings are unlikely to represent publication bias tilted toward a certain pattern of results (i.e., expected

Table 3
*Summary of Results From Meta-Analysis With Chinese and North American Data*

| Variable | Main effect | Interaction with age | Effect size[a] |
|---|---|---|---|
| Hong Kong vs. mainland China | $F(1, 146) = 5.26, p < .05$ | $F(1, 145) = 1.50, p > .20$ | 2.02 |
| Locales across North America and China | $F(3, 253) = 37.64, p < .001$ | $F(2, 251) = .58, p > .50$ | |
| U.S. vs. mainland China | | | 1.27 |
| Hong Kong vs. U.S. | | | 5.38 |
| Hong Kong vs. Canada | | | 13.21 |
| Locales across North America and China (Chinese conditions with think-falsely verbs) | $F(3, 189) = 22.49, p < .001$ | $F(2, 187) = .18, p > .50$ | |
| U.S. vs. mainland China | | | 1.55 |
| Hong Kong vs. U.S. | | | 3.52 |
| Hong Kong vs. Canada | | | 9.68 |

[a] Effect sizes are presented only for significant variables; effect sizes were computed as odds ratios.
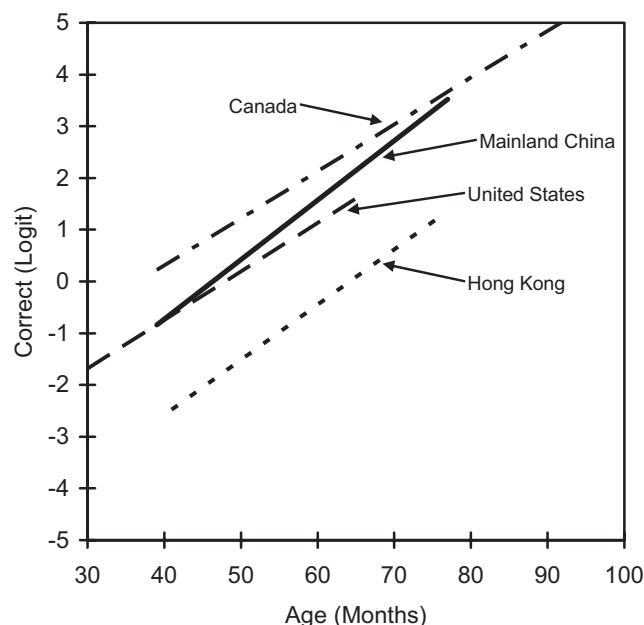
*Figure 2.* Developmental trajectories for the four communities— mainland China, Hong Kong, the United States, and Canada—of logit transformed data with increasing age.

similarity to prior research in Western European cultures). In addition, because the same pattern of findings was also found when only the data from published studies were analyzed, our findings are not skewed by the inclusion of unpublished data that have not undergone peer review.

First consider developmental trajectory. Chinese children develop from below- to above-chance performance during early childhood, and this parallels the trajectory of North American children (because there was not a significant interaction effect between locale and age in our results). Certainly, the developmental trajectory of Chinese versus North American children's false-belief understanding could have been nonparallel (because of differences in folk psychologies, differences in executive function development, the presence of think-falsely verbs in Chinese languages, and so on, as outlined in the Introduction).

None of the task variables we tested interacted significantly with age; variables showed either nonsignificance or simple main effects. With the nonsignificant findings, our data demonstrate that Chinese children's false-belief judgments are robust across many potentially relevant variables. Some have questioned whether the false-belief task always has validity in a non-Western context (e.g., Lillard, 1998). Here we show that not only is the trajectory the same in China as in Western cultures, but also it is important that similar task manipulations have the same non-effects (Wellman et al., 2001). This strengthens the case that false-belief tasks are measuring the same construct across cultures. Note also that the presence of think-falsely verbs in Chinese languages (and in many Chinese false-belief tasks) does not change the overall developmental trajectory for false-belief understanding; even with questions that use think-falsely verbs, children still develop from below- to above-chance performance.

The lack of an interaction for Chinese verb form with age undermines early competence accounts. Early competence ac-

counts of theory of mind (Chandler, Fritz, & Hala, 1989; Scholl & Leslie, 2001) have suggested that changes in false-belief performance during the preschool years reflect children's increasing ability to deal with verbal ambiguities of typical false-belief tasks or other domain-general cognitive demands of the tasks, such as executive functioning, rather than development in mental-state understanding. As such, if verbal ambiguities or domain-general cognitive demands were removed from false-belief tasks, false-belief performance would not improve with age during the preschool years, because 2- or 3-year-olds would already have the requisite mental-state understanding. Thus, the expected pattern in a meta-analysis for factors with different levels of verbal ambiguities would be an interaction with age on false-belief performance (Wellman et al., 2001). That is, for example, conditions that use the think-falsely verbs (which significantly lower verbal ambiguities) should show no (or considerably less) change with age in contrast to typical conditions. However, our results show that Chinese verb form did not interact with age. Furthermore, even in especially facilitative linguistic environments (i.e., Chinese children growing up with think-falsely verbs), understanding false belief poses genuine conceptual difficulties for young children and performance changes with age. It also appears that Chinese children have earlier competence at executive function tasks than North American children (Sabbagh et al., 2006; Tardif et al., 2007), although this does not translate into superior false-belief performance. Therefore, our pattern of findings undermines a variety of early competence accounts that contend that executive functioning limitations or linguistically demanding tasks have masked otherwise apparent early false-belief competence (see Sabbagh et al., 2006).

Next, consider developmental timing. Several variables—most clearly verb form and locale or language—visibly and significantly influenced the timing of false-belief understanding. Using think-falsely verbs in false-belief tasks improves the performance of Chinese children relative to using more neutral verbs. Two individual studies with Mandarin- (Lee et al., 1999) and Cantonese-speaking (Tardif et al., 2004) children also found an effect of using think-falsely verbs in the question. However, those studies did not compare Chinese children's performance to that of children in Western societies, as we were able to do in our analyses. As is clear in Figure 2, the presence of think-falsely verbs in Chinese languages does not enhance Chinese children's overall false-belief understanding relative to North American children.

It is most intriguing that the timing of development varies significantly across different locales, as shown in Figure 2. Such a large difference in timing between specific locales counters Callaghan et al.'s (2005) suggestion that the timing of false-belief understanding might be tightly synchronous across very different cultures. Instead, we find different developmental timetables exhibiting more than 2 years' difference across communities.

Although our results demonstrate significant differences in developmental timing across locales, the specific differences in timing (including the difference between mainland China and Hong Kong) admit no straightforward interpretations. Although there are numerous differences between the four populations that could lead to differences in false-belief performance, these differences do not easily account for our findings. For example, mainland Chinese children are almost certainly singletons, whereas the Hong Kong children are more likely to have siblings. Given research showing

that having older siblings is associated with enhanced false-belief understanding (Perner, Ruffman, & Leekam, 1994), one might expect Hong Kong's timetable to be closer to North American trajectories than to mainland China's timetable, but we found the opposite pattern of results. Likewise, because Hong Kong is more "Westernized" than mainland China (Bond & Cheung, 1983), one might expect a pattern of results opposite from the observed results. Also, children in Hong Kong are more likely to be bilingual (Tardif, Fletcher, Marchman, & Liang, 2006), a factor associated with enhanced false-belief understanding (Goetz, 2003), and yet their developmental timetable is the furthest behind.

It is also important to consider socioeconomic status (SES). The studies in our dataset contain little SES data; almost all the studies provide no such data. More importantly, we consider SES to be too broad a factor and unlikely to provide a simple explanation for the pattern of observed differences. It is unclear what it would mean to compare such a broad concept as SES in the different cultural contexts of the North American and the Chinese populations. For example, does having college-educated parents or being middle class mean the same thing in the cultural contexts of North America, mainland China, and Hong Kong? When we compare mainland China and Hong Kong, very little is known about the effect of SES on cognitive and language development. However, the little data that exists would suggest that SES has very little influence on comparisons of cognitive and language development between mainland China and Hong Kong. For example, in a norming study of the MacArthur-Bates Communicative Development Inventories with normative samples of 3,270 children from both Beijing and Hong Kong, the average educational levels of both parents (mothers and fathers) and both sets of grandparents (maternal and paternal) were all significantly lower for the Hong Kong sample than for the Beijing sample (Tardif et al., 2006). Nevertheless, the effect of Hong Kong versus Beijing on vocabulary score yielded an $R^2$ of only .03, and the effect of SES yielded an even smaller $R^2$ of .003. That is, there were only small differences between Hong Kong and Beijing in children's verbal ability, and SES played very little role in that difference. These findings suggest that the developmental timing differences observed between Hong Kong and mainland China (and thus between Hong Kong and North America) for false-belief understanding were not due to (or only very minimally influenced by) SES or, for that matter, verbal ability. Still, without direct data on SES or other possible differences between the populations between the locales, we cannot completely rule out the possibility that these variables play a larger role.

Generally, although one can speculate and point to numerous comparisons across the populations that could account for differences, it is noteworthy that no single sociocultural or linguistic factor (especially broad factors such as SES, East–West culture, or bilingualism) can provide a straightforward account of the pattern of differences observed in this meta-analysis. Instead, it seems more likely that a coalescence of factors as yet unknown will prove responsible. This counts as an important conclusion from the current research; it directs future research to examine specific variables (e.g., particular parental practices or linguistic features) to address the question of how specific adult folk psychologies, enculturation, and linguistic factors jointly shape the development of theory of mind.

## Conclusion

The current meta-analysis supports three main points. First, the results indicate that theory of mind universally develops. Specifically, Chinese and North American children all develop from below- to above-chance performance, and this trajectory does not disappear with the inclusion of any of the task variables we examined. Second, although it reflects parallel developmental trajectories, theory of mind understanding appears on substantially different timetables across numerous cultures and languages. Summing across these two points, the developmental course of theory of mind necessarily includes influences of both universal trajectories and specific experiential factors. Third, the observed timetable differences are unlikely to be explained by straightforward, simplistic accounts, but are the product of (and could be used to reveal) the multiple sociocultural and linguistic factors that jointly shape theory of mind development.

## References

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

*Astington, J. W., Gopnik, A., & O'Neill, D. (1989). *Young children's understanding of unfulfilled desire and false belief.* Unpublished manuscript, Ontario Institute for Studies in Education, Toronto, Ontario, Canada.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

*Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60,* 946–964.

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

Bond, M. H., & Cheung, T. S. (1983). College students' spontaneous self-concept: The effect of culture among respondents in Hong Kong, Japan, and the United States. *Journal of Cross-Cultural Psychology, 14,* 153–171.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., et al. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science, 16,* 378–384.

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

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

*Carpendale, J. L., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67,* 1686–1706.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small scale deceit: Deception as a marker of 2-, 3-, and 4-year-olds' early theories of mind. *Child Development, 60,* 1263–1277.

*Chandler, M., & Hala, S. (1994). The role of personal involvement in the assessment of early false belief skills. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind* (pp. 403–425). Hillsdale, NJ: Erlbaum.

*Chen, M. J., & Lin, Z. X. (1994). Chinese preschoolers' difficulty with theory-of-mind tests. *Bulletin of the Hong Kong Psychological Society, 32,* 34–46.

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., & Stewart, S. L. (1998). Child-rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: A cross-cultural study. *Developmental Psychology, 34,* 677–686.

*Dalke, D. E. (1995). Explaining young children's difficulty on the false

belief task: Representational deficits of context-sensitive knowledge? *British Journal of Developmental Psychology, 13,* 209–222.

*Davis, D. L. (1997, April). *Children's understanding of the role of knowledge and thinking in pretense.* Paper presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.

*Deng, C. (2001). *A study on development and representational mechanism of young children's theory of mind.* Unpublished doctoral dissertation, East China Normal University, Shanghai, China.

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.

*Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483–527.

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

*Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6,* 1–15.

*Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Development, 59,* 26–37.

*Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of 3-year-olds' understanding of false belief. *Child Development, 62,* 83–97.

*Hickling, A. K., Wellman, H. M., & Gottfried, G. (1997). Preschoolers' understanding of others' mental attitudes toward pretend happenings. *British Journal of Developmental Psychology, 15,* 339–354.

*Lee, K., Olson, D. R., & Torrance, N. (1999). Chinese children's understanding of false beliefs: The role of language. *Journal of Child Language, 26,* 1–21.

Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123,* 3–32.

*Lillard, A. S., & Flavell, J. H. (1992). Young children's understanding of different mental states. *Developmental Psychology, 28,* 626–634.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60,* 167–171.

*Moore, C., Pure, K., & Furrow, P. (1990). Children's understanding of the modal expression of certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development, 61,* 722–730.

*Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development, 8,* 1–25.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108,* 291–310.

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

Peterson, C. C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language, 15,* 123–145.

*Robinson, E. J., & Mitchell, P. (1995). Masking children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development, 66,* 1022–1039.

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

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73,* 734–751.

*Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory-of-mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17,* 74–81.

*Sang, B., Ma, L., & Deng, C. (2004). An experimental study between the using of mental terms and the development of theory of mind. *Psychological Science (China), 27,* 584–589.

Scholl, B. J., & Leslie, A. M. (2001). Minds, modules, and meta-analysis. *Child Development, 72,* 696–701.

*Sheffield, E. G., Sosa, B. B., & Hudson, J. A. (1993, March). *Understanding metarepresentation: 2- and 3-year-olds' comprehension of false belief.* Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.

Shweder, R. A., & Bourne, E. J. (1984). Does the concept of the person vary cross-culturally? In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 158–199). Cambridge, England: Cambridge University Press.

*Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understanding belief. *Child Development, 67,* 2967–2988.

*Sullivan, K., & Winner, E. (1991). When 3-year-olds understand ignorance, false belief and representational change. *British Journal of Developmental Psychology, 9,* 159–171.

*Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56,* 135–148.

Tardif, T. (2006). But are they really verbs? Chinese words for action. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 477–498). New York: Oxford University Press.

Tardif, T., Fletcher, P., Marchman, V., & Liang, W. (2006, June). *Comprehension and production of nouns and verbs: Data from the CDI norming studies in English, Mandarin, and Cantonese.* Paper presented at the International Conference on Infant Studies, Kyoto, Japan.

*Tardif, T., Fung, K., Wellman, H. M., Liu, D., & Fang, F. (2001, April). *Preschoolers' understanding of knowing how, knowing that, and false belief.* Poster session presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

*Tardif, T., & Ng, M. C. (2001). *Emotional understanding and theory of mind development.* Unpublished manuscript, Chinese University of Hong Kong, China.

*Tardif, T., So, C., & Kaciroti, N. (2007). The development of false belief and its relation to syntactic complements in Cantonese-speaking preschoolers. *Developmental Psychology, 43,* 318–340.

*Tardif, T., & Wellman, H. M. (2000). Acquisition of mental state language in Mandarin- and Cantonese-speaking children. *Developmental Psychology, 36,* 25–43.

*Tardif, T., Wellman, H. M., & Cheung, K. M. (2004). False belief understanding in Cantonese-speaking children. *Journal of Child Language, 31,* 779–800.

*Tardif, T., Yang, X. D., & So, C. (2000). *False belief and complement understanding in Mandarin.* Unpublished manuscript, Chinese University of Hong Kong, China.

*Taylor, M., & Carlson, S. M. (1997). The relation between individual differences in fantasy and theory of mind. *Child Development, 68,* 436–455.

*Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Vinden, P. G. (1999). Children's understanding of mind and emotion: A multi-cultural study. *Cognition and Emotion, 13,* 19–48.

*Volling, B. L., Xiang, Z., & Niu, W. (2001). *Children's false belief understanding in Chinese and U.S. preschoolers.* Unpublished manuscript, University of Michigan, Ann Arbor.

Wang, Q. (2004). The emergence of cultural self-construct: Autobiographical memory and self-description in American and Chinese children. *Developmental Psychology, 40,* 3–15.

Wellman, H. M. (1998). Culture, variation and levels of analysis in our folk psychologies. *Psychological Bulletin, 123,* 33–36.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684.

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.

*Winner, E., & Sullivan, K. (1993). *Deception as a zone of proximal development for false belief understanding.* Unpublished manuscript, Boston College.

*Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition, 35,* 41–68.