Developmental Neuropsychology

ISSN: 8756-5641 (Print) 1532-6942 (Online) Journal homepage: www.tandfonline.com/journals/hdvn20

# Age-Related Changes in Theory of Mind in Middle Childhood: A Cross-Cultural Comparison Between Australian and Chinese Children

Si Wang, Glenda Andrews, Donna Pendergast, David L. Neumann, Xiao Liang, Yuan Cao, Duo Li & David H.K. Shum

**To cite this article:** Si Wang, Glenda Andrews, Donna Pendergast, David L. Neumann, Xiao Liang, Yuan Cao, Duo Li & David H.K. Shum (15 Sep 2025): Age-Related Changes in Theory of Mind in Middle Childhood: A Cross-Cultural Comparison Between Australian and Chinese Children, Developmental Neuropsychology, DOI: 10.1080/87565641.2025.2560482

**To link to this article:** https://doi.org/10.1080/87565641.2025.2560482



Published online: 15 Sep 2025.

DEVELOPMENTAL NEUROPSYCHOLOGY
https://doi.org/10.1080/87565641.2025.2560482



# Age-Related Changes in Theory of Mind in Middle Childhood: A Cross-Cultural Comparison Between Australian and Chinese Children

Si Wang[a], Glenda Andrews[a], Donna Pendergast[b], David L. Neumann[a], Xiao Liang[c,d], Yuan Cao[e], Duo Li[c], and David H.K. Shum[c,d]

[a]School of Applied Psychology, Griffith University, Brisbane, Queensland, Australia; [b]School of Education and Professional Studies, Griffith University, Brisbane, Queensland, Australia; [c]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, SAR, China; [d]Mental Health Research Centre, The Hong Kong Polytechnic University, Hong Kong, SAR, China; [e]Department of Social Work and Social Administration, University of Hong Kong, Hong Kong, SAR, China

**ABSTRACT**

Theory of mind (ToM) is centrally important in everyday social communication and interactions, and a growing number of studies have focused on this social-cognitive construct in school-aged children. This study explored age-related changes in ToM abilities and cross-cultural differences between children from China and Australia. We recruited 126 children from China and 83 children from Australia. The children's cognitive and affective ToM were measured by four tests. The findings showed that although there were no cultural differences in Interpretive ToM and Faux Pas performances, Chinese children performed poorer in Reading Mind in the Eyes and Strange Stories than Australian children.

## Introduction

Theory of mind (ToM) refers to the ability to attribute mental states to oneself and others (Premack & Woodruff, 1978). This construct develops between 3 and 5 years of age (Wellman et al., 2011). Children develop social cognition through triadic interaction with different social contexts and the child's environment (e.g., family environment, educational settings) (Carpendale & Lewis, 2004). For instance, research has shown that educational environments can significantly influence ToM development in middle school children (Smogorzewska et al., 2020). Children begin to build social connections outside of the family, such as peer relationships, which could facilitate their ToM (Slaughter et al., 2002). The change in social context is a good point of view for researching ToM in middle childhood (Osterhaus & Koerber, 2021), especially from a cross-cultural perspective. Given that ToM is a foundational social-cognitive skill that influences children's functioning (Carlson et al., 2013), cross-cultural comparisons are necessary because sociocultural factors play important roles in children's ToM (Z. Wang et al., 2016).

### *Advanced theory of mind in middle childhood*

Advanced ToM is a broad concept and has been classified in different ways. Earlier research (Shamay-Tsoory et al., 2005) classified ToM measures according to the level of emotional demand they incorporated, namely, cognitive ToM, affective ToM, and mixed tasks. Cognitive ToM refers to the

**CONTACT** David H.K. Shum david.shum@polyu.edu.hk Mental Health Research Centre, The Hong The Hong Kong Polytechnic University, A401a, Chung Sze Yuen Building, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong, SAR, China

ability to make inferences about beliefs and motivations, while affective ToM refers to the ability to infer what a person is feeling (Sebastian et al., 2011). The mixed tasks assess both cognitive and affective ToM. A more recent conceptual framework proposed by Osterhaus et al. (2016) identified three dimensions of ToM, namely, social reasoning, recognizing transgression of social norms, and reasoning about ambiguity. Social reasoning involves understanding and interpreting others' mental states in social contexts and can be measured by the Reading Mind in the Eyes (RMIE; Baron-Cohen et al., 2001) and Strange Stories (Happé, 1994). Recognizing transgression of social norms involves detecting when social norms are violated and can be measured by the Faux Pas test (Banerjee & Watling, 2005). Reasoning about ambiguity pertains to navigating situations where information is unclear or uncertain and can be measured by the Interpretive Theory of Mind task (Lalonde & Chandler, 2002). This three-dimensional structure of ToM was supported by the results of recent studies (Osterhaus & Koerber, 2021; Osterhaus et al., 2016). Specifically, in Osterhaus et al.'s study Osterhaus et al. (2016), a series of ToM tasks, including false belief tasks, Strange Stories, RMIE, Faux Pas test, and ambiguity tasks, were administered to elementary school children (8-, 9-, and 10-year-olds). Exploratory and confirmatory factor analyses revealed these three distinct ToM factors.

Middle childhood is a crucial period for social and cognitive development, characterized by significant advancements in ToM (Osterhaus & Bosacki, 2022; Weimer et al., 2021). Both cross-sectional and longitudinal studies have provided evidence for the developments of ToM in middle childhood. For example, Lagattuta et al. (2010) examined both 4- to 9-year-old children and adults' performance on the Interpretive Theory of Mind (IToM) task. They demonstrated that the ability to appreciate diversity and common ground in beliefs continued to develop into adulthood. Cantin et al. (2016) found that age directly predicted children's performance on the Strange Stories test in a sample of 93 children, aged 7 to 10 years old. Wilson et al. (2018) also found age-related development in 5- to 12-year-old Australian children's performance on the Strange Stories test. Osterhaus et al. (2016) found significant age-related development for social reasoning as measured by the Strange Stories test and reasoning about ambiguity as measured by IToM task, but not for recognizing transgressions of social norms as measured by Faux Pas task in 8- to 10-year-old children.

Longitudinal studies have been conducted recently and provided more direct evidence of the developmental trajectories of different dimensions of ToM. Lecce et al. (2017) reported growth in performance on the Strange Stories test based on a longitudinal design by following 113 children from 9.5 years to 10.5 years of age. Smogorzewska et al. (2024) found that ToM assessed by the six-step ToM scale and false belief task continued to develop from 7.5 years to 9.5 years of age. Osterhaus et al. (2024) found longitudinal associations between reading comprehension and social reasoning (RMIE, Strange Stories, and false belief tasks), but not between reading comprehension and recognition of social norm transgression (Faux Pas) in 112 nine-year-olds over 1 year. Results of another two longitudinal studies indicated that recognizing transgressions of social norms likely reaches a plateau phase in middle elementary school, around 7 years (Osterhaus & Koerber, 2021; Smogorzewska & Osterhaus, 2021). To summarize, different dimensions of ToM develop in different trajectories. As children develop their ToM skills, they improve their ability to manage peer relationships and understand social norms (Osterhaus et al., 2025). This improved capability to navigate social situations allows them to interpret social cues accurately and respond appropriately, resulting in more successful interactions with their peers (Lane & Bowman, 2021). The development of ToM is crucial for school-age children. These skills form the foundation for effective communication and collaboration, which are essential in both academic and social contexts (Lecce & Devine, 2022; Smogorzewska et al., 2024).

## *Cultural comparison of the theory of mind in middle childhood*

There are many reasons ToM might differ between Western and Eastern cultures. A recent meta-analysis (Yu & Wellman, 2024) compared the developmental trajectory of ToM across 22 countries and found 2- to 10-year-old children from individualistic and collectivist countries had different sequences shaped by cultural learning. Specifically, the ability of children from individualistic

countries in understanding people's beliefs could differ before understanding basic aspects of knowledge acquisition. However, children from collectivist countries understood basic aspects of knowledge acquisition before they understood the diversity of beliefs. This cultural difference in the development of ToM may be related to the Confucius culture, which has long prioritized attaining the common knowledge that all people should know. Individualist culture, on the other hand, emphasizes more on understanding beliefs and belief divergence: Individuals have the right to hold their own beliefs. Chinese children learn to pay attention to external factors (e.g., rules, guidance), and self-control is encouraged. In contrast, children raised in most Western countries learn more about personal autonomy and expressing individual opinions and mental states (Slaughter & Perez-Zapata, 2014). These cultural factors may affect the processing and understanding of one's own emotions and those of others (i.e., social reasoning). Previous studies' findings suggest the presence of cultural differences in ToM in early childhood. For example, although children from both North America and China passed the false-belief task at around 4 to 5 years old, children from Hong Kong and Japan showed a 2-year lag on false-belief understanding (D. Liu et al., 2008, Naito & Koyama, 2006). In another study, Chinese children passed a knowledge ignorance task before they passed the diverse beliefs task (Wellman et al., 2006), whereas the order of acquisition was reversed in US and Australian children (Shahaeian et al., 2011; Wellman et al., 2011). One recent study compared ToM and inhibitory control in 3–6-year-old children between Japan and Scotland and found that inhibitory control and metacognition predict ToM competence in Scottish children rather than in Japanese children, supporting the Western developmental route to ToM (Symeonidou et al., 2022). Similarly, Fujita et al. (2022) found 3–6-year-old children in Japan showed worse ToM performance than a matched sample in UK. These findings might reflect the value placed on pragmatic knowledge in Eastern cultures and on personal ideas and beliefs in Western cultures.

There is some evidence for cultural differences in ToM for school-aged-children. Kobayashi et al. (2007) examined the neural correlates of second-order false-belief understanding among 8- to 12-year-old Japanese/English bilingual and English-speaking monolingual children. The fMRI findings showed that the monolingual group had stronger activations in the right inferior parietal lobe and the overlapping temporoparietal junction, whereas the bilingual group had stronger activations in the left superior temporal sulcus and the overlapping temporal pole. A recent study directly compared school-aged children from the United Kingdom and Hong Kong, and found that the children in Hong Kong performed poorer on ToM tasks (Z. Wang et al., 2016). However, Hong Kong may not be typical regarding traditional Chinese culture (Bond & Cheung, 1983). Therefore, the question of whether there is any difference in ToM between mainland Chinese and Western school-aged children remains unanswered.

### *Aims and hypotheses*

In this study, we aimed to assess age-related differences in the three dimensions of ToM (viz., social reasoning, recognizing transgression of social norms, and reasoning about ambiguity) in school-aged children and to compare cultural differences between Australian and mainland Chinese children. Understanding cross-cultural differences in the ToM is essential for advancing both theoretical knowledge and practical applications in developmental psychology. Firstly, examining these differences helps researchers identify which aspects of ToM are universal and which are influenced by cultural contexts. This distinction is vital for grasping how various cultures shape the development and expression of understanding mental states and for challenging assumptions predominantly based on Western research. Secondly, exploring cross-cultural differences enriches our models of ToM development. It emphasizes how culture shapes social cognition and illustrates the flexibility of social-cognitive development across diverse cultural contexts. Thirdly, investigating cultural differences in ToM highlights the importance of employing culturally sensitive research methods. This approach ensures that findings are accurate and relevant across varied cultural contexts. We used a battery of four commonly used and age-appropriate tests of ToM to comprehensively examine children's ToM: IToM (reasoning about

ambiguity), Faux Pas (recognizing transgression of social norms), Reading Mind in the Eyes (RMIE) (social reasoning), and Strange Stories (social reasoning) tests. For IToM and RMIE, these are replicable across cultures and have low reliability on verbal responses. Faux Pas and Strange Stories consist of culturally appropriate scenarios suitable for a wide age range in middle childhood. All tasks have been used in both cultures and are considered adapted for both contexts (S. Wang et al., 2021).

Based on previous research, we hypothesized that age-related differences would appear in the three dimensions of ToM in both cultural groups (Lagattuta et al., 2010; Lecce et al., 2017; Wilson et al., 2018). Given that the timing of grasping false-belief understanding is equivalent in preschoolers from mainland China and North America (D. Liu et al., 2008), we hypothesized that both groups of children would show parallel age-related differences in performance on the reasoning about ambiguity (i.e., IToM), which (like false belief) focuses on understanding another's thoughts and beliefs. In contrast, because Chinese culture emphasizes self-control and emotional constraint (Markus & Kitayama, 1991; Matsumoto et al., 2008), school-aged Chinese children may express personal emotions less frequently. Therefore, we hypothesized that Chinese children would perform worse than their Australian counterparts on the Faux Pas, which requires recognizing when someone says something socially inappropriate or hurtful. Similarly, for the tasks involving broader social reasoning, RMIE and Strange Stories, we hypothesized that Chinese children would perform worse than their Australian counterparts. These tasks require not only inferring others' emotions and intentions but also drawing on a broader set of social-cognitive inferences, which may be influenced by cultural norms about emotional expressivity and social interpretation (Fu et al., 2023).

## Method

### *Participants*

Participants were recruited from one primary school in Brisbane, a city with a population of 2.51 million in 2020 (Australian Bureau of Statistics, 2020), and one school in Xinyang, a city in central China with a population of 6.47 million in 2019 (Henan Province Bureau of Statistics, 2019). Xinyang is less developed and has had less exposure to Western cultures. It is famous for its "tea culture," which emphasizes harmony and respect, and is closely linked to traditional Chinese culture and ethics (Yang, 2007). Given this background, the Xinyang sample can be considered a more traditional and less Westernized group than the Beijing and Hong Kong samples used in previous studies were Z. Wang et al. (2016) and Wellman et al. (2006). Australians and Chinese were selected for this study because they have been previously compared as representatives of Western and Eastern cultures (Chen et al., 2015; Hiew et al., 2015).

The Australian sample was recruited from a predominantly middle-class community in Brisbane, while the Chinese sample was recruited from a middle-size city (Xinyang) in Central China, which is an area representing Chinese traditional culture through its rich historical heritage where archeological evidence is dated to the Neolithic period (L. Liu & Chen, 2012) and active cultural preservation which cumulated through cultural development through major dynasties (Zhang, 2018).

To recruit participants, information sheets and consent forms were sent home with students via their schools. Parents provided informed consent and additional information to ensure their children met the eligibility criteria.

The 83 Australian children (43 boys and 40 girls) and 126 Chinese children (74 boys and 52 girls) all met the following inclusion criteria:

(1) Australian participants born and raised in Australia; Chinese participants born and raised in mainland China;
(2) For Australian participants, English as the first language; for Chinese participants, Chinese (Mandarin) as the first language;
(3) Normal or corrected-to-normal vision and hearing;

(4) No reported history of brain injury, premature birth, or diagnosis of behavioral or learning disorders.

The participants ($N = 209$) were grouped according to age. Group 1 (G1) included all children aged from 5 years, 6 months to 7 years, 11 months; Group 2 (G2) included all children aged from 8 years to 9 years, 11 months; Group 3 (G3) included all children aged from 10 years to 11 years, 11 months (see Table 1). Independent samples *t*-tests showed that there were no significant differences in age between the Australian and Chinese children (all $ps > .05$).

The parent's education levels are presented in Table 2. Chi-square tests showed that culture was significantly associated with the mother's ($\chi^2$ (5, $N = 209$) = 57.56, $p < .001$, Cramer's $V = .53$) and father's education level ($\chi^2$ (5, $N = 207$) = 56.84, $p < .001$, Cramer's $V = .52$). The Australian mothers and fathers had higher levels of education than the Chinese mothers and fathers.

## Measures and procedure

The child participants completed the IToM, Faux Pas, RMIE, and Strange Stories ToM tasks and the Peabody Picture Vocabulary Test. For the Chinese versions of the IToM, Faux Pas and Strange Stories measures, we adopted a back-translation approach for all scripts and coding guidelines (Brislin, 1970). Three Chinese/English bilingual developmental psychology researchers (one PhD candidate, two psychologists with a PhD degree) translated each word into Chinese and then back-translated into English. For the RMIE, the Autism Research Centre at University of Cambridge provided a simplified Chinese version of the RMIE-C. The same images were used in the English and Chinese versions. A pilot study was used to confirm that Chinese children could understand the descriptors of all ToM tasks.

### Reasoning about ambiguity: IToM

Lalonde and Chandler's task (2002) were used to assess the children's interpretive understanding of others' minds. The participants had to interpret ambiguous parts of drawings in displays from the perspectives of two cartoon characters. The displays were presented using PowerPoint software. There

**Table 1.** Numbers and mean ages of children in each age group and culture.

| | Chinese | | | | Australian | | | |
|---|---|---|---|---|---|---|---|---|
| Gender | G1 | G2 | G3 | Total | G1 | G2 | G3 | Total |
| Male (*n*) | 23 | 26 | 25 | 74 | 20 | 14 | 9 | 43 |
| Female (*n*) | 23 | 16 | 13 | 52 | 11 | 16 | 13 | 40 |
| Total (*n*) | 46 | 42 | 38 | 126 | 31 | 30 | 22 | 83 |
| $M_{age}$ in years | 6.93 | 8.97 | 10.70 | 8.75 | 6.65 | 9.05 | 10.94 | 8.65 |
| (SD) | (0.64) | (0.61) | (0.43) | (1.65) | (0.59) | (0.59) | (0.55) | (1.82) |

*Note*. For ease of reference, we have labeled the age groups as follows in this paper: G1 = 5:6–7:11 years. G2 = 8:0–9:11 years. G3: 10:0–11:11 years.

**Table 2.** Frequencies of parents with each level of education by culture.

| | Chinese | | Australian | |
|---|---|---|---|---|
| Education | Mother ($N = 126$) | Father ($N = 126$) | Mother ($N = 83$) | Father ($N = 81$) |
| Year 6 and less | 0 (0.0%) | 1 (0.8%) | 3 (3.6%) | 2 (2.4%) |
| Year 10 | 12 (9.5%) | 12 (9.5%) | 3 (3.6%) | 6 (7.2%) |
| Year 12 | 17 (13.5%) | 17 (13.5%) | 5 (6.0%) | 12 (14.5%) |
| Some University | 32 (25.4%) | 38 (30.2%) | 8 (9.6%) | 6 (7.2%) |
| Bachelor | 64 (50.8%) | 58 (46.0%) | 35 (42.2%) | 28 (33.7%) |
| Postgraduate | 1 (0.8%) | 0 (0.0%) | 29 (34.9%) | 27 (32.5%) |

*Note*. Education level was coded as 1 = year 6 and less, 2 = year 10, 3 = year 12, 4 = some university, 5 = bachelor, and 6 = postgraduate (masters and above).

were four trials of the task. Trials 1 and 2 used a drawing of a ship arriving too late to save a drowning witch; Trials 3 and 4 used a drawing of an elephant and a grapefruit. The cartoon characters (Tim and Jane) were located either inside or outside a house. The experimenter ensured the participant understood that Tim and Jane could not hear or see anything outside the house unless they were invited out. For Trials 1 and 2, Slide 1 displayed the entire original picture, "a ship arriving too late to save a drowning witch." After discussing the picture with the participant, a blue shape flew in and covered the picture with a small rectangular cutout that formed a small viewing window. The cutout revealed only the part of the picture showing the ship's bow and the witch's pointed hat. Thus, the visible part only showed two triangles. Slide 2 (Trial 1) showed Tim on the left side of the screen. The participant was told that Tim had never seen the picture before nor heard any part of the discussion of what the full drawing depicted. Then, the experimenter asked the test question, "What does Tim think this is?" After the participant answered, Slide 3 was presented (Trial 2). Tim did not appear on Slide 3. Jane was invited out, and she appeared on the right side of the screen. The question was then repeated for her. The same procedure was carried out for the second picture, "an elephant and a grapefruit" (Trials 3 and 4) but the sequence and order of the appearance of Tim and Jane were reversed.

We used a coding procedure based on Lalonde and Chandler's study (2002). Responses that reflected a mistaken understanding that the characters could see what the entire pictures entailed were coded as reality errors. For example, in Trial 1, a child said the cartoon character thinks that the triangle in the cutoff part is a ship. Less explicit mistakes that contained trace elements from the full picture were scored as contamination errors. For example, for Trial 1, a child said the cartoon character thinks the triangle is a sail. If the participant provided the same attribution for the two characters (e.g., "both Jane and Tim think it is a mountain"), it was scored as an identical response. The total number of reality and contamination errors ranged from 0 to 4, and the total number of identical responses ranged from 0 to 2.

An interpretive response for one picture was coded when there was no reality/contamination error, and the participant made different attributions for the two characters. For example, for Trial 1, a child said that Tim thinks the cutout shows a cliff and a tree, while Jane thinks it is the wing of a bird and the roof of a house. Participants who gave interpretive responses for both pictures received a score of 2 (interpretive). Participants who gave interpretive responses for just one picture received a score of 1 (transitional). Participants who gave two noninterpretive responses received a score of 0 (noninterpretive). The total score for IToM ranged from 0 to 2.

### *Recognizing transgressions of social norms: Faux Pas*

We chose the two faux pas stories from Banerjee and Watling's study (2005). The participants heard two different stories, and they were asked some questions after each story. Each story involved two cartoon characters. One character said something without considering that the utterance might hurt the other character. Six forced-choice questions tested the children's understanding of faux pas. They included detection (did someone say something they should not have said?), identification (what was said that should not have been said?), feelings (how does [insulted character] feel now?), intention (did [insulting character] want to make [insulted character] upset?), comprehension (question regarding target object), and ignorance (did [insulting character] know [insulted character's relationship with the target object]?). The presentation order of correct and incorrect response options was counterbalanced between stories. The understanding of each story was scored as 1 if the participant answered all six questions correctly. The total score for the Faux Pas task ranged from 0 to 2. Considering there were two options for each of the six questions, the probability of getting full marks for a story by guessing was .016. For each of the six faux pas question types, a score out of 2 was calculated to represent the number of stories in which the participant answered the question type correctly.

### *Social reasoning: reading the mind in the eyes (RMIE)*

The RMIE test was used to assess the ability to read emotions in another's eyes (Baron-Cohen et al., 2001). We used the child version of RMIE (RMIE-C), which comprises a series of 28 pictures of the eye

region of the faces of different people. The child was asked to use visual information to choose a word that best described the mental state term that he/she was feeling or thinking based on the pictures. For each item of the test, a picture is framed with four words (one target state, three distractor states) that describe the possible mental state term of the person presented in the picture. The number of correct choices was recorded. The final score ranged from 0 to 28.

### *Social reasoning: Strange Stories*

A subset of the 12 short vignettes of the Strange Stories task was used to assess the development of advanced ToM (Happé, 1994; O'Hare et al., 2009). The version of Strange Stories we used in the current study has been cross-culturally validated with children from mainland China and Australia in our previous study (S. Wang et al., 2021). The stories included 12 scenarios: a lie, a white lie, joking, pretending, misunderstanding, persuasion, appearance reality, a figure of speech, sarcasm, forgetting, a double bluff, and contrary emotions. Each story presented a character who, within the context of the particular scenario, said something untrue. The participants were asked, "Why did she/he say this?" to assess ToM. Responses were scored from 0 to 2, with 0 indicating an incorrect or physical state response, 1 indicating a partial psychological state response, and 2 indicating a full and accurate psychological state response. The final scores ranged from 0 to 24.

### *Verbal ability measures: peabody picture vocabulary test (PPVT)*

To control for the effect of verbal ability on ToM, we used the PPVT-4 (Dunn & Dunn, 2007). The English version was employed with Australian children. The split-half reliability for Form A and Form B ranged from .90 to .97 for ages 5–11 years. We used the Chinese version of the PPVT-Revised Edition (PPVT-R) with the Chinese children (Sang & Miao, 1990). Studies have shown that the test has excellent psychometric properties. Because different versions of the PPVT were administered to the two cultural samples, we standardized and used the PPVT raw scores for both the Chinese and the Australian children in the analyses.

## Data analysis

The measures derived from the IToM and Faux Pas tasks were categorical. The age group and cultural differences were analyzed using chi-square tests. Cramer's $V$ determined the effect size and it was interpreted based on Gravetter and Wallnau's criteria (Gravetter & Wallnau, 2004). If a significant effect of culture or age group was found, ordinal regression was conducted to verify the significance with verbal ability controlled.

The measures derived from the RMIE and Strange Stories tasks were continuous. Responses were analyzed using 2 (Culture: Chinese vs. Australian) × 3 (Age group: G1 vs. G2 vs. G3) ANCOVAs with verbal ability controlled. The effect size for the ANCOVA was determined using a partial eta square (Cohen, 1988).

This study was not preregistered. The data set is not publicly available to protect the participants' confidentiality, but it is available from the first author upon a reasonable request.

## Results

### *Correlation between demographics, verbal ability, and ToM*

For both Chinese and Australian children, performance on all ToM tasks positively correlated with age and verbal ability ($ps < .001$; Table 3). Parents' education levels and gender did not correlate with the children's performance on any ToM task in either cultural group ($ps > .05$). Consequently, we used verbal ability as a control variable in subsequent analyses. There were significant correlations among the four ToM tasks within each cultural group. Partial correlation analyses showed that after

**Table 3.** Correlation among demographics, verbal ability, and ToM tasks in Chinese (above diagonal) and Australian children (below diagonal).

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1.Age (years) | – | .79*** | −.04 | −.02 | −.09 | .36*** | .43*** | .27** | .58*** |
| 2.Verbal ability | .79*** | – | .05 | .03 | −.10 | .35*** | .44*** | .32*** | .64*** |
| 3. Mother's Edu | .01 | .16 | – | .61*** | −.07 | .09 | −.02 | .14 | .06 |
| 4. Father's Edu | −.03 | .11 | .59*** | – | −.06 | −.07 | −.02 | −.02 | −.08 |
| 5. Gender | .18 | −.01 | .07 | −.14 | – | −.10 | −.03 | .09 | .06 |
| 6. IToM | .30** | .30** | −.01 | −.05 | .04 | – | .19* | .17 | .19* |
| 7. Faux Pas | .44*** | .47*** | .05 | −.01 | .07 | .06 | – | .29** | .34*** |
| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 8. RMIE | .48*** | .42*** | .17 | .02 | .13 | .07 | .32** | – | .46*** |
| 9. Strange Stories | .63*** | .65*** | .16 | .10 | .18 | .28* | .36** | .25* | – |

*Note.* *$p < .05$.**$p < .01$.***$p < .001$.Verbal ability was represented by standardized PPVT raw scores; Mother's Edu = Mother's education level; Father's Edu = Father's education level, Gender was coded as 0 = male, 1 = female.

controlling for age and verbal ability, one correlation between RMIE and Strange Stories in the Chinese sample ($r = .35$, $p < .001$) remained significant (Table 4).

### *Age-group and culture differences in IToM*

A three-way log-linear analysis verified that the three-way interaction among IToM response category, culture, and age group interaction was not significant. The final model the log-linear analysis produced only retained the IToM response category × age group interaction. The likelihood ratio of this model was $\chi^2$ (8) = 7.87, $p = .45$. From this, we performed two chi-square tests of independence to examine the relation between culture and IToM response category and the relation between age group and IToM response category, respectively. Table 5 shows the percentages of children in each response category within each age and cultural group. The association between IToM response category and culture was not significant, $\chi^2$ (2, $N = 209$) = 2.19, $p = .33$, Cramer's $V = .10$. In contrast, the association of IToM response category with age group was significant, $\chi^2$ (4, $N = 209$) = 26.91, $p < .001$, Cramer's $V = .25$. A post hoc test showed that more children in G1 were coded as non-interpretive and fewer children in G1 were coded as interpretive, compared with expected counts. In contrast, fewer children in G3 were coded as non-interpretive and more children in G3 were coded as interpretive, compared with the expected counts.

**Table 4.** Partial correlation among demographics and ToM tasks in Chinese (above diagonal) and Australian children (below diagonal) after controlling for age and verbal ability.

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Mother's Edu | – | .60*** | .07 | −.10 | .03 | −.14 | −.06 |
| 2. Father's Edu | .57*** | – | .06 | .07 | .02 | .03 | .12 |
| 3. Gender | −.13 | .09 | – | −.06 | .02 | .13 | .17 |
| 4. IToM | .04 | .07 | −.01 | – | .02 | .06 | −.07 |
| 5. Faux Pas | .02 | .05 | .04 | −.12 | – | .18 | .06 |
| 6. RMIE | −.17 | −.02 | .05 | −.11 | .11 | – | .35*** |
| 7. Strange Stories | −.08 | .10 | .15 | .09 | .03 | −.13 | – |

*Note.* *$p < .05$.**$p < .01$.***$p < .001$.Verbal ability was represented by standardized PPVT raw scores; Mother's Edu = Mother's education level; Father's Edu = Father's education level, Gender was coded as 0 = male, 1 = female.

**Table 5.** Percentage (%) of participants in each score of IToM by age group across cultures.

| | Chinese ($n = 126$) | | | | Australian ($n = 83$) | | | |
|---|---|---|---|---|---|---|---|---|
| Score | G1 | G2 | G3 | All | G1 | G2 | G3 | All |
| 0 | 52.2 | 31.0 | 13.2 | 33.3 | 67.7 | 26.7 | 31.8 | 43.4 |
| 1 | 30.4 | 42.9 | 34.2 | 35.7 | 19.4 | 36.7 | 40.9 | 31.3 |
| 2 | 17.4 | 26.2 | 52.6 | 31.0 | 12.9 | 36.7 | 27.3 | 25.3 |

*Note.* Score 0 = noninterpretive. Score 1 = transitional. Score 2 = interpretive. G1 = 5:6–7:11; G2 = 8:0–9:11; G3: 10:0–11:11.

We conducted a further ordinal regression analysis to investigate the effect of age group on IToM response category in the combined samples, controlling for verbal ability. Age was significant in the model. For each 1-year increase in age, the odds of scoring 2 on the IToM task versus 0 or 1 combined were 1.28 times greater (95% CI, 1.00–1.63), $B = .25$, $SE = .13$, Wald $\chi^2$ (1) = 3.88, $p = .049$. Verbal ability was not significant in the model, $B = .33$, $SE = .22$, Wald $\chi^2$ (1) = 2.37, $p = .12$, odds ratio = 1.40 (95% CI, 0.91–2.14).

To investigate the errors children made on the IToM task in detail, we analyzed their reality errors, contamination errors, and identical responses.

We used two 3 (Age: G1 vs. G2 vs. G3) × 2 (Culture: Chinese vs. Australian) ANCOVAs with verbal ability controlled to examine the main and interactive effects of the two independent variables on children's reality/contamination errors and identical responses in the IToM task. For reality/contamination errors, the assumption of homogeneity of variance was not met. After log transforming the dependent variable (reality/contamination errors) to restore the equal variance, the results' pattern did not change. Therefore, we used untransformed data in the analysis. The main effect of culture was significant, $F(1, 202) = 5.27$, $p = .02$, partial $\eta^2 = .03$, indicating that the Chinese children made more reality or contamination errors than the Australian children did. The main effect of age group was not significant, $F(2, 202) = 1.26$, $p = .29$, partial $\eta^2 = .01$. No significant interaction between culture and age group was found, $F(2, 202) = 0.15$, $p = .86$, partial $\eta^2 = .001$. The assumption of homogeneity of variance was also not met for identical responses; however, the appropriate data transformation did not change the results' pattern. The main effect of culture was significant, $F(1, 202) = 8.08$, $p = .005$, partial $\eta^2 = .04$, Australian children reported more identical responses than the Chinese children did. The main effect of age group was not significant, $F(2, 202) = 1.75$, $p = .18$, partial $\eta^2 = .02$, but there was a significant interaction between culture and age group, $F(2, 202) = 3.51$, $p = .03$, partial $\eta^2 = .03$. Simple main effect analyses showed no significant difference in identical responses between Chinese and Australian children in G1 and G2 ($ps > .05$). However, in G3, the Australian children made more identical responses than the Chinese children did ($p = .001$) (see Table 6).

In sum, the Chinese and Australian children did not differ significantly in their IToM scores when age and verbal ability were controlled, but there were cultural differences in the types of errors made. The Chinese children made more reality errors than the Australian children did. The Australian children in G3 made more identical responses than their Chinese counterparts did. Meanwhile, significant age differences in IToM scores were revealed after controlling for verbal ability. The children in G1 had lower IToM scores compared to those in G2 and G3.

### *Age-group and culture differences in the Faux Pas test*

We performed two chi-square tests of independence to examine how the Faux Pas score related to culture and age group (see Table 7). The relationship between the Faux Pas score and culture was not significant, $\chi^2$ (2, $N = 209$) = 0.523, $p = .770$, Cramer's $V = .05$. The Chinese children's performance was similar to that of the Australian children. In contrast, the relationship with age group was significant, $\chi^2$ (4, $N = 209$) = 44.97, $p < .001$, Cramer's $V = .33$. A post hoc test showed that G1 contained more children with scores of 0 and fewer children with scores of 2 compared with expected counts. In contrast, fewer children in G3 scored 0 and more scored 2 compared to expected counts.

We conducted ordinal regression analysis to investigate the effect of age on the Faux Pas score using the combined data set, controlling for verbal ability. Age was not significant in the

**Table 6.** Means and standard deviations of types of errors in IToM by age group across cultures.

| | Chinese ($n = 126$) | | | | Australian ($n = 83$) | | | |
|---|---|---|---|---|---|---|---|---|
| Error Type | G1 | G2 | G3 | All | G1 | G2 | G3 | All |
| Reality/Contamination | 1.39 (1.42) | 0.71 (0.89) | 0.47 (0.73) | 0.89 (1.14) | 1.10 (1.35) | 0.37 (0.61) | 0.09 (0.29) | 0.57 (1.00) |
| Identical responses | 0.96 (0.84) | 0.74 (0.77) | 0.24 (0.54) | 0.67 (0.79) | 1.19 (0.87) | 0.73 (0.87) | 0.95 (0.84) | 0.96 (0.88) |

**Table 7.** Percentage (%) of participants in each score of Faux Pas by age group across cultures.

| | Chinese ($n = 126$) | | | | Australian ($n = 83$) | | | |
|---|---|---|---|---|---|---|---|---|
| Score | G1 | G2 | G3 | All | G1 | G2 | G3 | All |
| 0 | 65.2 | 16.7 | 21.1 | 35.7 | 54.8 | 33.3 | 4.5 | 33.7 |
| 1 | 21.7 | 45.2 | 28.9 | 31.7 | 29.0 | 30.0 | 27.3 | 28.9 |
| 2 | 13.0 | 38.1 | 50.0 | 32.5 | 16.1 | 36.7 | 68.2 | 37.3 |

*Note*. Score 0 = answered some questions incorrectly in both stories. Score 1 = answered all questions correctly only in one story. Score 2 = answered all questions correctly in two stories.

model, $B = .21$, $SE = .13$, Wald $\chi^2$ (1) = 2.73, $p = .10$, odds ratio = 1.24 (95% CI, 0.96–1.60). Only verbal ability contributed to the model. For every 1 unit increase in the standardized PPVT score, the odds of scoring 2 on the Faux Pas test versus 0 and 1 combined were 2.09 times greater (95% CI, 1.31–3.35), $B = .74$, $SE = .24$, Wald $\chi^2$ (1) = 9.47, $p = .002$. In sum, the Chinese children's Faux Pas scores were similar to the Australian children's scores. Age was not a significant predictor of performance on the Faux Pas test after controlling for verbal ability.

Ceiling effects appeared in the children's responses to the detection (pass rate > 84%; Table 8) and feeling questions (pass rate > 96%) in both cultural groups. Consistent with Banerjee and Watling (2005) and Banerjee et al.'s findings Banerjee et al. (2011) the intention and ignorance questions were answered less accurately than the other questions were.

Next, we performed ordinal regression analyses to examine how culture, age, and verbal ability predicted the children's responses to the other faux pas questions. Culture was the only significant predictor of faux pas identification. Compared to the Australian children, the odds of scoring 2 versus 0 and 1 combined in the Chinese sample were 0.323 times smaller (95% CI, 0.13–0.83), $B = -1.13$, $SE = .48$, Wald $\chi^2$ (1) = 5.49, $p = .02$. Neither age nor verbal ability was significant. This result indicates that more Australian children than Chinese children answered the identification questions correctly. Culture, age, and verbal ability were all nonsignificant for both faux pas intention and faux pas comprehension.

For faux pas ignorance, both age and verbal ability were significant predictors, but culture was nonsignificant. For every 1 year increase in age, the odds of scoring 2 on the ignorance question versus 0 and 1 combined were 1.30 times greater (95% CI, 1.00–1.68), $B = .26$, $SE = .13$, Wald $\chi^2$ (1) = 3.86, $p = .05$. For every 1 unit increase in the standardized PPVT score, the odds of scoring 2 versus 0 and 1 combined were 1.80 times greater (95% CI, 1.14–2.83), $B = .59$, $SE = .23$, Wald $\chi^2$ (1) = 6.40, $p = .01$.

In conclusion, cultural differences were only found in the children's responses to the identification questions. The Chinese children were less accurate than the Australian children were on identification questions. Differences related to age and verbal ability were only revealed in the children's responses to ignorance questions, which tested whether the children understood the false belief held by the person who committed the faux pas.

### *Age-group and culture differences in RMIE*

We used a 2 × 3 two-way ANCOVA to investigate the relationships of culture (Chinese vs. Australian) and age group (G1 vs. G2 vs. G3) with the children's performance on RMIE after controlling for verbal ability (see Table 9). The assumption of homogeneity of variance was met. The main effect of culture was significant, $F(1, 202) = 55.26$, $p < .001$, partial $\eta^2 = .22$, indicating the Chinese children ($M = 16.17$, $SD = 2.85$) performed worse on RMIE than their Australian counterparts did ($M = 18.93$, $SD = 2.84$). The main effect of age group was not significant, $F(2, 202) = 0.86$, $p = .43$, partial $\eta^2 = .01$. There was no significant interaction between culture and age group, $F(2, 202) = 0.88$, $p = .42$, partial $\eta^2 = .01$.

### *Age-group and culture differences on the Strange Stories Task*

A two-way ANCOVA was used to investigate how culture (Chinese vs. Australian) and age group (G1 vs. G2 vs. G3) related to Strange Stories performance, controlling for verbal ability (see Table 10). The

**Table 8.** Percentage (%) of participants in each score of each Faux Pas question by age group across cultures.

| | Chinese ($n = 126$) | | | | Australian ($n = 83$) | | | |
|---|---|---|---|---|---|---|---|---|
| Score | G1 | G2 | G3 | All | G1 | G2 | G3 | All |
| Detection | | | | | | | | |
| 0 | 2.2 | 4.8 | 7.9 | 4.8 | 16.1 | 6.7 | 0.0 | 8.4 |
| 1 | 6.5 | 11.9 | 10.5 | 9.5 | 12.9 | 6.7 | 0.0 | 7.2 |
| 2 | 91.3 | 83.3 | 81.6 | 85.7 | 71.0 | 86.7 | 100.0 | 84.3 |
| Identification | | | | | | | | |
| 0 | 6.5 | 2.4 | 5.3 | 4.8 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1 | 13.0 | 9.5 | 21.1 | 14.3 | 12.9 | 3.3 | 4.5 | 7.2 |
| 2 | 80.4 | 88.1 | 73.7 | 81.0 | 87.1 | 96.7 | 95.5 | 92.8 |
| Feelings | | | | | | | | |
| 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1 | 6.5 | 2.4 | 0.0 | 3.2 | 6.5 | 3.3 | 0.0 | 3.6 |
| 2 | 93.5 | 97.6 | 100.0 | 96.8 | 93.5 | 96.7 | 100.0 | 96.4 |
| Intention | | | | | | | | |
| 0 | 17.4 | 2.4 | 0.0 | 7.1 | 12.9 | 6.7 | 4.5 | 8.4 |
| 1 | 23.9 | 14.3 | 15.8 | 18.3 | 19.4 | 13.3 | 18.2 | 16.9 |
| 2 | 58.7 | 83.3 | 84.2 | 74.6 | 67.7 | 80.0 | 77.3 | 74.7 |
| Comprehension | | | | | | | | |
| 0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.7 | 3.3 | 0.0 | 4.8 |
| 1 | 19.6 | 2.4 | 7.9 | 10.3 | 12.9 | 16.7 | 9.1 | 13.3 |
| 2 | 80.4 | 97.6 | 92.1 | 89.7 | 77.4 | 80.0 | 90.9 | 81.9 |

| Score | Chinese ($n = 126$) | | | | Australian ($n = 83$) | | | |
|---|---|---|---|---|---|---|---|---|
| | G1 | G2 | G3 | All | G1 | G2 | G3 | All |
| Ignorance | | | | | | | | |
| 0 | 30.4 | 9.5 | 5.3 | 15.9 | 22.6 | 13.3 | 0.0 | 13.3 |
| 1 | 50.0 | 45.2 | 13.2 | 37.3 | 32.3 | 40.0 | 13.6 | 30.1 |
| 2 | 19.6 | 45.2 | 81.6 | 46.8 | 45.2 | 46.7 | 86.4 | 56.6 |

**Table 9.** Means and standard deviations for RMIE (raw scores) by age group across cultures.

| | Chinese | | | Australian | | |
|---|---|---|---|---|---|---|
| Age group | *N* | *M* | *SD* | *N* | *M* | *SD* |
| G1 | 46 | 15.30 | 3.05 | 31 | 17.61 | 2.69 |
| G2 | 42 | 16.36 | 2.53 | 30 | 19.03 | 2.71 |
| G3 | 38 | 17.00 | 2.73 | 22 | 20.64 | 2.34 |
| All | 126 | 16.17 | 2.85 | 83 | 18.93 | 2.84 |

*Note*. The possible score of RMIE ranges from 0 to 28.

**Table 10.** Means and standard deviations for Strange Stories (raw scores) by age group across cultures.

| | Chinese | | | Australian | | |
|---|---|---|---|---|---|---|
| Age group | *N* | *M* | *SD* | *N* | *M* | *SD* |
| G1 | 46 | 11.33 | 4.19 | 31 | 12.84 | 3.09 |
| G2 | 42 | 14.95 | 3.12 | 30 | 16.87 | 3.44 |
| G3 | 38 | 16.66 | 2.88 | 22 | 19.05 | 2.74 |
| All | 126 | 14.14 | 4.13 | 83 | 15.94 | 4.01 |

*Note*. The possible score of Strange Stories ranges from 0 to 24.

assumption of homogeneity of variance was not met, but the appropriate data transformation did not change the pattern of results. The main effect of culture was significant, $F(1, 202) = 17.47$, $p < .001$, partial $\eta^2 = .08$, indicating that the Chinese children ($M = 14.14$, $SD = 4.13$) performed worse on the Strange Stories Task than their Australian counterparts did ($M = 15.94$, $SD = 4.02$). The main effect of age group was also significant, $F(2, 202) = 5.28$, $p = .006$, partial $\eta^2 = .05$. Post hoc comparisons using the Bonferroni test indicated that the mean score of G1 ($M = 11.94$, $SD = 3.84$) was significantly lower than that of the other age groups (G2: $M = 15.75$, $SD = 3.37$; G3: $M = 17.53$, $SD = 3.03$). We found no significant interaction between culture and age group, $F(2, 202) = 0.49$, $p = .62$, partial $\eta^2 = .01$.

## Discussion

This study provides new evidence about ToM development in middle childhood (viz., 5.5- to 12-year-old Chinese and Australian children).

### *Age, culture, and IToM*

IToM requires children to recognize that different people can have diverse beliefs (Lagattuta et al., 2013) and diverse interpretations of the same situation (Chandler & Helm, 1984). Consistent with our hypothesis, we found no difference in IToM scores between Chinese and Australian children. We also found that age, but not verbal ability, predicted children's IToM (Hayward & Homer, 2017; Kennedy et al., 2015). These results indicate that IToM develops through middle childhood, and its performance may be less dependent on verbal ability than other ToM tasks.

Although we found no cultural difference in IToM scores, the types of errors differed between the cultural groups. The Chinese children made more reality/contamination errors than the Australian children did a possible explanation of this is that the Chinese children might have used their analogical thinking to draw on prior knowledge ("the circle is a grapefruit") to predict the naïve person's belief ("the character thinks the circle is an orange"), based on the same-shape relation between grapefruit and oranges. A previous cross-cultural study showed that Hong Kong preschoolers used analogical reasoning and outperformed US preschoolers on processing complex relations (Richland et al., 2010). The authors noted that the advantage in complex analogies might be due to Chinese children's greater experience with socialized relational inputs. From this, Chinese children's better analogical thinking might increase their reality/contamination errors.

Older Australian children (10:0–11:11) made more identical responses than older Chinese children. Identical responses indicate that the children predicted that both naïve characters had the same belief about the restricted view of the picture. The result for the Australian sample was consistent with previous findings in US children and adults. Lagattuta et al. (2013) reported that identical responses increased with age and that 8- to 9-year-old US older children and adults most frequently provided them, while reality/contamination errors decreased with age. In our study, the Australian children might have inferred that the two characters had the same knowledge about the same object based on their daily experiences with siblings or friends. The collectivistic culture in China values interpersonal relationship and group functioning (Bochner, 1994; Markus & Kitayama, 1991). From this, Chinese children are educated to pay more attention to others'

perspectives; thus, it may be easier for them to attribute different thoughts about the same reality to different people.

### Age, culture, and Faux Pas

Contrary to expectation, we found no cultural differences between the Chinese and Australian children's total Faux Pas scores. A recent study found that preschoolers from China and Western Europe showed similar trends of emotion comprehension, but the Chinese children performed better on recognizing hidden emotions (Tang et al., 2017).

The children in G3 (10:0–11:11) performed better than those in G1 (5:6–7:11) on the Faux Pas task. This is consistent with previous studies using the same tasks, which reported significant age-related differences in Faux Pas without considering the role of verbal ability (Banerjee & Watling, 2005; Banerjee et al., 2011). However, we found that age did not significantly predict Faux Pas after controlling for verbal ability, which means that children's development in verbal ability accounts for the variance in Faux Pas during middle childhood. This might be due to the high verbal demand this task imposed.

Our data did not reveal any cultural or age-related differences in the detection, feeling, intention, or comprehension questions. We only found cultural differences for the identification questions. The Australian children performed better on identifying which utterance should not be said. The reason why the Chinese children answered incorrectly is perhaps due to their over-interpretation of the scenario based on their real-life interaction with friends, rather than an inability to identify faux pas. Some Chinese children who chose the wrong answer explained that the character whose feelings had been hurt should understand that the other child did not say it intentionally and should tolerate the offense because they were good friends. Interestingly, Australian children who answered incorrectly never volunteered this kind of explanation. The specific explanation Chinese children provided aligned with the Chinese culture, which praises the sacrifice of personal interest for maintaining group harmony.

Among the six questions, the ignorance question was the most difficult. It tested whether the children could recognize that the faux pas was a consequence of a false belief. The age-related differences in school-aged children's responses to the ignorance question might relate to the development of executive functions such as working memory and inhibition control (Banerjee et al., 2011; Bottiroli et al., 2016).

### Age, culture, and RMIE

Overall, the Australian children outperformed the Chinese children on RMIE, consistent with our hypotheses. Given the emphasis on emotional constraint in Chinese culture (Markus & Kitayama, 1991; Matsumoto et al., 2008), school-aged Chinese children may have a lower frequency of exposure to and expression of complex personal emotions. Furthermore, Chinese preschoolers exhibit less emotion knowledge than Euro-American children (Q. Wang, 2008). These factors might have contributed to the Chinese children's poorer performance on RMIE compared with the Australian children. However, another reason may be that the RMIE used in our study was based on photos of Caucasian people's eyes and has been predominantly used in Western samples. Previous cross-cultural comparisons of adults' RMIE performance suggest that participants performed better on same-culture stimuli than on other-culture stimuli (Adams et al., 2010; Bjornsdottir & Rule, 2016). This indicates an intracultural advantage in attributing others' mental states. From this, the cultural differences revealed in our two school-aged samples may be due to the Australian children having an own-culture advantage.

Surprisingly, although the children's RMIE performance increased with age, there were no age-group differences after controlling for verbal ability. This suggests that the individual differences in RMIE largely relied on the children's knowledge of mental-state words. The RMIE requires children to

understand the meaning of the target words, which relates to vocabulary knowledge. Data from a college sample also showed that verbal IQ accounted for almost 25% of the variance in RMIE performance (E. Peterson & Miller, 2012), confirming the importance of verbal ability for RMIE performance.

### Age, culture, and Strange Stories

We found both cultural and age-related differences in the children's performance on Strange Stories, which aligned with our hypotheses. First, the Australian children outperformed their Chinese counterparts on Strange Stories, which was consistent with our previous study (S. Wang et al., 2021). Because the Strange Stories measure consists of social scenarios with various affects, our findings reflect contrasts in the relative salience of various mental states across cultures. The cultural difference in Strange Stories performance might reflect a difference in cultural values in which individuals' external behaviors are valued higher in Chinese culture, while personal internal mental states are suppressed (Y. Liu et al., 2014; Markus & Kitayama, 1991). Additionally, the cultural difference may hinge on the use of mental-state verbs to gauge the children's interpretation rather than their comprehension of others' emotional states. As noted above, Chinese preschoolers showed a similar trend of emotion comprehension to children from Western Europe, but they performed better in recognizing hidden emotions (Tang et al., 2017). However, we adopted the coding guidelines for Strange Stories from O'Hare et al. (2009), and it required the children to mention the exact concept or mental verbs to obtain a full score. Previous studies found that Chinese mothers' behavioral talk contributes to their children's false-belief understanding in early childhood (Y. Liu et al., 2014; Lu et al., 2008), whereas mothers' mental state talk strongly correlates with Western children's ToM (Adrian et al., 2005; C. Peterson & Slaughter, 2003). Presumably, talking about behavior to Chinese children might be more common than mentioning mental states. From this, the Chinese sample may have understood the mental states involved in the stories but used behavioral descriptions rather than mental concepts or verbs to explain the characters' responses.

Second, we found that the children aged 5.5 to 7 years old performed worse than the children aged above 8 years did after controlling for verbal ability. This indicates that children's flexibility in comprehending others' mental states in complex social situations improves significantly after age 7 years in middle childhood. It also suggests that younger and older school-aged children might differ in terms of the cognitive resources needed for processing advanced ToM.

### Contribution and implications

First, unlike most previous studies that used one or two tasks to capture ToM in middle childhood (Banerjee et al., 2011; Bock et al., 2014; Hayward & Homer, 2017), the current study employed four age-appropriate tests to examine three dimensions of ToM (viz., reasoning about ambiguity, recognizing transgression of social norms, and social reasoning) in two school-aged samples of children from China and Australia. Second, our findings suggest a distinction among the three dimensions of ToM in that cultural differences were observed in social reasoning (RMIE and Strange Stories), but not in reasoning about ambiguity (IToM) or in recognizing transgression of social norms (Faux Pas). Therefore, it is not enough to use one or two measures to represent the entire conceptual framework of school-aged children's ToM. Researchers should consider the specific ToM component measured in each task when selecting measures and interpreting their findings. Third, the current research highlights that cultural differences exist in some aspects of ToM and research indicates that whether ToM is a culturally universal or culturally specific construct might depend on the ToM facet measured in each task. Besides, although Chinese and Australian children performed at the same accuracy level of IToM, cultural differences were revealed in the types of errors they made in the task. This indicates that future cross-cultural research should examine both the

quantitative (e.g., accuracy) and qualitative (e.g., types of error) aspects of ToM performance to clarify their findings.

### *Limitations and future research*

The current study is not without limitations. We recruited a relatively small sample of Australian participants. Future studies with larger samples are needed. Additionally, the Chinese version of the Strange Stories task was directly translated from the English version that O'Hare et al. (2009) used. Similarly, the RMIE involved interpreting Caucasian faces. Therefore, the Chinese children's poorer performance might relate to their lowered familiarity with some social scenarios involved in the Strange Stories task. From this, it is advisable to adapt the stories to Chinese children's idiomatic expressions and life experiences. Furthermore, other factors, such as sibling composition variables and parenting practices, may underpin some of the observed cultural differences. Including these and other factors in future cross-cultural research will further the understanding of the social and cultural contributors to children's ToM development during the school years.

## Conclusion

This study's findings suggest that ToM is not a unidimensional but a multi-component construct in school-aged children. Different age-related and cultural differences were revealed for the four measures. Generally, younger children aged from 5.5 to 7 years performed worse than older children who were aged from 8 to 11 years. However, after controlling for verbal ability, age significantly predicted performance on the IToM and Strange Stories but not the Faux Pas and RMIE tests. There were no cultural differences in the overall scores of children's IToM and Faux Pas. However, a detailed analysis of errors reflected cultural influences on the tasks. In contrast, Australian children outperformed Chinese children on RMIE and Strange Stories. The findings could be attributed to either the superiority of Australian children's social reasoning due to their Western cultural background or the own-culture stimulus advantage in RMIE and Strange Stories tests. Further studies are needed to investigate what types of cultural factors may contribute to the individual differences in ToM in Chinese and Australian school-aged children.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Griffith University Postgraduate Scholarship and Yeung Tsang Wing Yee and Tsang Wing Hing Endowed Professorship in Neuropsychology in The Hong Kong Polytechnic University.

## ORCID

Xiao Liang http://orcid.org/0000-0002-9874-3376

## Availability of data and materials

The dataset is not publicly available to protect the participants' confidentiality but is available from the first author on reasonable request.

## References


Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97–108. https://doi.org/10.1162/jocn.2009.21187

Adrian, J. E., Clemente, R. A., Villanueva, L., & Rieffe, C. (2005). Parent-child picture-book reading, mothers' mental state language and children's theory of mind. *Journal of Child Language*, *32*(3), 673–686. https://doi.org/10.1017/s0305000905006963

Australian Bureau of Statistics, Statistics About The Population and Components Of Change (Births, Deaths, Migration) For Australia's Capital Cities And Regions. (2020). Retrieved April, 2024, from https://www.abs.gov.au/statistics/people/population/regional-population/latest-release#capital-cities

Banerjee, R., & Watling, D. (2005). Children's understanding of faux pas: Association with peer relationship. *Hellenic Journal of Psychology*, *2*, 27–45. *Special Issue on Theory of Mind*. https://pure.royalholloway.ac.uk/ws/portalfiles/portal/1414152/Banerjee_Watling_05.pdf

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, *82*(6), 1887–1905. https://doi.org/10.1111/j.1467-8624.2011.01669.x

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Bjornsdottir, R. T., & Rule, N. O. (2016). On the relationship between acculturation and intercultural understanding: Insight from the reading the mind in the eyes test. *International Journal of Intercultural Relations*, *52*, 39–48. https://doi.org/10.1016/j.ijintrel.2016.03.003

Bochner, S. (1994). Cross-cultural differences in the self concept. *Journal of Cross-Cultural Psychology*, *25*(2), 273–283. https://doi.org/10.1177/0022022194252007

Bock, A. M., Gallaway, K. C., & Hund, A. M. (2014). Specifying links between executive functioning and theory of mind during middle childhood: Cognitive flexibility predicts social understanding. *Journal of Cognition and Development*, *16*(3), 509–521. https://doi.org/10.1080/15248372.2014.888350

Bond, M. H., & Cheung, T.-S. (1983). College students' spontaneous self-concept. *Journal of Cross-Cultural Psychology*, *14*(2), 153–171. https://doi.org/10.1177/0022002183014002002

Bottiroli, S., Cavallini, E., Ceccato, I., Vecchi, T., & Lecce, S. (2016). Theory of mind in aging: Comparing cognitive and affective components in the faux pas test. *Archives of Gerontology & Geriatrics*, *62*, 152–162. https://doi.org/10.1016/j.archger.2015.09.009

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson McInnis, M. S., & Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *Journal of Experimental Child Psychology*, *146*, 66–78. https://doi.org/10.1016/j.jecp.2016.01.014

Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *WIREs Cognitive Science*, *4*(4), 391–402. https://doi.org/10.1002/wcs.1232

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(1), 79–96. https://doi.org/10.1017/S0140525X04000032

Chandler, M. J., & Helm, D. (1984). Developmental changes in the contribution of shared experience to social role-taking competence. *International Journal of Behavioral Development*, *7*(2), 145–156. https://doi.org/10.1177/016502548400700203

Chen, X. J., Liu, L. L., Cui, J. F., Wang, Y., Shum, D. H., & Chan, R. C. (2015). Chinese and Australians showed differences in mental time travel in emotion and content but not specificity. *Frontiers in Psychology*, *6*, 879. https://doi.org/10.3389/fpsyg.2015.00879

Cohen, J. (1988). *Statiscal power analysis for the behavioral sciences* (2nd ed.). La Wrence Erlabaum Associates.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test* (4th ed.). Pearson.

Fu, I. N., Chen, K. L., Liu, M. R., Jiang, D. R., Hsieh, C. L., & Lee, S. C. (2023). A systematic review of measures of theory of mind for children. *Developmental Review*, *67*, 101061. https://doi.org/10.1016/j.dr.2022.101061

Fujita, N., Devine, R. T., & Hughes, C. (2022). Theory of mind and executive function in early childhood: A cross-cultural investigation. *Cognitive Development*, *61*, 101150. https://doi.org/10.1016/j.cogdev.2021.101150

Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the behavioral sciences* (6th ed.), Wadsworth Cengage Learning.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism & Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/bf02172093

Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, *35*(3), 454–462. https://doi.org/10.1111/bjdp.12186

Henan Province Bureau of Statistics, 2019 Xinyang city national economic and social development statistical bulletin (in Chinese). Retrieved April, 2024, from http://tjj.xinyang.gov.cn/www/tjzl/tjgb/2020/0429/26107.html
Hiew, D. N., Halford, W. K., Van de Vijver, F. J., & Liu, S. (2015). Relationship standards and satisfaction in Chinese, western, and intercultural Chinese-western couples in Australia. *Journal of Cross-Cultural Psychology*, *46*(5), 684–701. https://doi.org/10.1177/0022022115579936
Kennedy, K., Lagattuta, K. H., & Sayfan, L. (2015). Sibling composition, executive function, and children's thinking about mental diversity. *Journal of Experimental Child Psychology*, *132*, 121–139. https://doi.org/10.1016/j.jecp.2014.11.007
Kobayashi, C., Glover, G. H., & Temple, E. (2007). Cultural and linguistic effects on neural bases of 'Theory of Mind' in American and Japanese children. *Brain Research*, *1164*, 95–107. https://doi.org/10.1016/j.brainres.2007.06.022
Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, *46*(6), 1417–1432. https://doi.org/10.1037/a0021062
Lagattuta, K. H., Sayfan, L., & Harvey, C. (2013). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development*, *85*(2), 659–674. https://doi.org/10.1111/cdev.12154
Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology*, *20*(2–3), 163–198. https://doi.org/10.1016/s0732-118x(02)00007-7
Lane, J. D., & Bowman, L. C. (2021). How children's social tendencies can shape their theory of mind development: Access and attention to social information. *Developmental Review*, *61*, 100977. https://doi.org/10.1016/j.dr.2021.100977
Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology*, *163*, 69–86. https://doi.org/10.1016/j.jecp.2017.06.011
Lecce, S., & Devine, R. T. (2022). Theory of mind at school: Academic outcomes and the influence of the school context. *Infant and Child Development*, *31*(1), e2274. https://doi.org/10.1002/icd.2274
Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*(2), 523–531. https://doi.org/10.1037/0012-1649.44.2.523
Liu, L., & Chen, X. (2012). *The archaeology of China: From the late Paleolithic to the early Bronze Age*. Cambridge University Press.
Liu, Y., Wang, Y., Luo, R., & Su, Y. (2014). From the external to the internal: Behavior clarifications facilitate theory of mind (ToM) development in Chinese children. *International Journal of Behavioral Development*, *40*(1), 21–30. https://doi.org/10.1177/0165025414562484
Lu, H., Su, Y., & Wang, Q. (2008). Talking about others facilitates theory of mind in Chinese preschoolers. *Developmental Psychology*, *44*(6), 1726–1736. https://doi.org/10.1037/a0013074
Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295x.98.2.224
Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality & Social Psychology*, *94*(6), 925–937. https://doi.org/10.1037/0022-3514.94.6.925
Naito, M., & Koyama, K. (2006). The development of false-belief understanding in Japanese children: Delay and difference? *International Journal of Behavioral Development*, *30*(4), 290–304. https://doi.org/10.1177/0165025406063622
O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism & Developmental Disorders*, *39*(6), 916–928. https://doi.org/10.1007/s10803-009-0699-2
Osterhaus, C., & Bosacki, S. L. (2022). Looking for the lighthouse: A systematic review of advanced theory-of-mind tests beyond preschool. *Developmental Review*, *64*, 101021. https://doi.org/10.1016/j.dr.2022.101021
Osterhaus, C., D'Urso, G., Koerber, S., & Bosacki, S. L. (2025). Theory of mind, self-perceptions, and peer popularity in middle childhood and early adolescence. *Children*, *12*(3), 281. https://doi.org/10.3390/children12030281
Osterhaus, C., & Koerber, S. (2021). The development of advanced theory of mind in middle childhood: A longitudinal study from age 5 to 10 years. *Child Development*, *92*(5), 1872–1888. https://doi.org/10.1111/cdev.13627
Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child Development*, *87*(6), 1971–1991. https://doi.org/10.1111/cdev.12566
Osterhaus, C., Lecce, S., & Koerber, S. (2024). Unlocking narratives: Longitudinal associations between theory of mind and reading comprehension. *British Journal of Developmental Psychology*, *42*(4), 511–516. https://doi.org/10.1111/bjdp.12514
Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, *18*(3), 399–429. https://doi.org/10.1016/s0885-2014(03)00041-8
Peterson, E., & Miller, S. F. (2012). The eyes test as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00220

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Richland, L. E., Chan, T.-K., Morrison, R. G., & Au, T. K.-F. (2010). Young children's analogical reasoning across cultures: Similarities and differences. *Journal of Experimental Child Psychology*, *105*(1–2), 146–153. https://doi.org/10.1016/j.jecp.2009.08.003

Sang, B., & Miao, X. (1990). The revision of trail norm of Peabody picture vocabulary test revised (PPVT-R) in Shanghai proper. *Psychological Science (Chinese Journal)*, *5*, 20–25. https://caod.oriprobe.com/articles/37645961/pi_bo_di_tu_pian_ci_hui_ce_yan_xiu_ding_ban__ppvt_.htm

Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S.-J., De Brito, S. A., McCrory, E. J., & Viding, E. (2011). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*(1), 53–63. https://doi.org/10.1093/scan/nsr023

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239–1247. https://doi.org/10.1037/a0023899

Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, *18*(1), 55–67. https://doi.org/10.1097/01.wnn.0000152228.90129.99

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology*, *20*(4), 545–564. https://doi.org/10.1348/026151002760390945

Slaughter, V., & Perez-Zapata, D. (2014). Cultural variations in the development of mind reading. *Child Development Perspectives*, *8*(4), 237–241. https://doi.org/10.1111/cdep.12091

Smogorzewska, J., & Osterhaus, C. (2021). Advanced theory of mind in children with mild intellectual disability and deaf or hard of hearing children: A two-year longitudinal study in middle childhood. *British Journal of Developmental Psychology*, *39*(4), 603–624. https://doi.org/10.1111/bjdp.12389

Smogorzewska, J., Szumski, G., Bosacki, S., Grygiel, P., & Osterhaus, C. (2024). Longitudinal relations between theory of mind and academic achievement among deaf and hard-of-hearing school-aged children. *Journal of Experimental Child Psychology*, *239*, 105806. https://doi.org/10.1016/j.jecp.2023.105806

Smogorzewska, J., Szumski, G., Grygiel, P., & Wang, Z. (2020). Theory of mind goes to school: Does educational environment influence the development of theory of mind in middle childhood? *PLOS ONE*, *15*(8), e0237524. https://doi.org/10.1371/journal.pone.0237524

Symeonidou, M., Mizokawa, A., Kabaya, S., Doherty, M., & Ross, J. (2022). Contrasting one's share of the shared life space: Comparing the roles of metacognition and inhibitory control in the development of theory of mind among Scottish and Japanese children Developmental Science 27 5 e13417. https://doi.org/10.31234/osf.io/wj6kx

Tang, Y., Harris, P. L., Pons, F., Zou, H., Zhang, W., & Xu, Q. (2017). The understanding of emotion among young Chinese children. *International Journal of Behavioral Development*, *42*(5), 512–517. https://doi.org/10.1177/0165025417741366

Wang, Q. (2008). Emotion knowledge and autobiographical memory across the preschool years: A cross-cultural longitudinal investigation. *Cognition*, *108*(1), 117–135. https://doi.org/10.1016/j.cognition.2008.02.002

Wang, S., Andrews, G., Pendergast, D., Neumann, D., Chen, Y., & Shum, D. H. (2021). A cross-cultural study of theory of mind using strange stories in school-aged children from Australia and mainland China. *Journal of Cognition and Development*, *23*(1), 40–63. https://doi.org/10.1080/15248372.2021.1974445

Wang, Z., Devine, R. T., Wong, K. K., & Hughes, C. (2016). Theory of mind and executive function during middle childhood across cultures. *Journal of Experimental Child Psychology*, *149*, 6–22. https://doi.org/10.1016/j.jecp.2015.09.028

Weimer, A. A., Warnell, K. R., Ettekal, I., Cartwright, K. B., Guajardo, N. R., & Liew, J. (2021). Correlates and antecedents of theory of mind development during middle childhood and adolescence: An integrated model. *Developmental Review*, *59*, 100945. https://doi.org/10.1016/j.dr.2020.100945

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, *17*(12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, *82*(3), 780–792. https://doi.org/10.1111/j.1467-8624.2011.01583.x

Wilson, J., Andrews, G., Hogan, C., Wang, S., & Shum, D. H. (2018). Executive function in middle childhood and the relationship with theory of mind. *Developmental Neuropsychology*, *43*(3), 163–182. https://doi.org/10.1080/87565641.2018.1440296

Yang, Z. (2007). Tea culture and Sino-American tea connections. *Chinese American Forum*. April 2024, http://caforumonline.net/CAFHandlerPDF.ashx?ID=208

Yu, C. L., & Wellman, H. M. (2024). A meta-analysis of sequences in theory-of-mind understandings: Theory of mind scale findings across different cultural contexts. *Developmental Review*, *74*, 101162. https://doi.org/10.1016/j.dr.2024.101162

Zhang, M. (2018). Research on the inheritance and development of Xinyang traditional culture. *Henan Social Science*, *26*(3), 45–52. in Chinese.