THEORY OF MIND IN OLD ADULTS: 
THE PERFORMANCE ON HAPPÉ’S STORIES AND FAUX PAS STORIES 

Yifang WANG\textsuperscript{1,2} and Yanjie SU\textsuperscript{1,2} 
\textsuperscript{1)Peking University, China}  
\textsuperscript{2)Capital Normal University, China} 

The research reported here was supported by grants from NSFC (30370201), Ministry of Education of China (01JAXLX014) and Beijing Key Laboratory to Su Yanjie. We are also grateful to Prof. Ellen Winner (Boston College, U.S.A) for her helpful comments on earlier drafts of this article. Correspondence concerning this article should be addressed to Su Yanjie, Department of Psychology, Peking University, Beijing, China, 100871; E-mail: yjsu@pku.edu.cn, Tel.: 86-10-62751833, Fax: 86-10-62756460.

Theory of mind is the ability to ascribe mental states to explain behavior (Leslie, 1987; Premack & Woodruff, 1978). In the past decades, developmental scientists focused on false belief understanding acquired at the age of 4–6 years (e.g., Sullivan, Zaitchik, & Tager-Flusberg, 1994; Wimmer & Perner, 1983). However, the development of theory of mind abilities beyond the age of 4–6 years old attracted growing interest in recent years. Happé (1994) found that children did not understand “strange stories” until the age of 8 involving lie, white lie, joke, pretense, misunderstanding, persuasion, appearance/reality, figure of speech, sarcasm, forgetting, double bluff, and contrary emotions. And Baron-Cohen, O’Riordan, Stone, Jones, and Plaisted (1999) found that 11 year olds could well recognize “faux pas”, which is whether individuals can recognize when someone unintentionally says something that would hurt or insult the other (Stone, Baron-Cohen, & Knight, 1998).

Moreover, Happé, Winner, and Brownell (1998) did a pioneering examination of theory of mind (measured by materials adapted from Happé, 1994) to explore the possible differences between the performances of young and old adults. Concerning questions of...
strange stories that required the ability to attribute mental states to the characters in the story, the performance of the old participants (mean age = 73 years) was superior to the one of the young participants (mean age = 21 years). However, concerning questions of control stories that did not require the ability of mental state understanding, the performance of the old participants did not differ significantly from that of the young participants. Their results seem to tell us that theory of mind abilities develop and improve throughout the life. The authors gave several possible explanations for this finding: (1) There is a “dedicated processing mechanism for theory of mind” (p. 360). (2) The old participants were more intelligent than the young participants because the old group was from Boston University School of Medicine and therefore well-educated ($M = 14$ years 7 months). Hence, Happé et al. (1998) suggested that “a study including IQ measures would be necessary” (p. 360). Maylor, Moulson, Muncer, and Taylor (2002) and Sullivan and Ruffman (2004) replicated and extended the work of Happé et al., with crystallized and fluid intelligence measures included. Many stories in the two studies were based on Happé (1994) and Happé et al. However, the contrary result was found that theory of mind abilities in the old group were worse than those in the young group. Further, Sullivan and Ruffman demonstrated the worse performance of the old participants on two new theory of mind tasks, identifying emotions from still photos and video clips and identifying cognitions from video clips.

The first purpose of the present study was to explore the possible differences between old and young adults concerning the performance on strange stories (theory of mind tasks), faux pas stories (theory of mind tasks), and physical stories (control tasks), with full-scale IQ (measured by Wechsler Adult Intelligence Scale-Revised, WAIS-R) and educational level matched.

The second purpose was to see whether the developmental trends from youth to old age concerning strange stories and faux pas stories were different. Research showed that the ability to understand strange stories develops earlier than the ability to recognize faux pas (Baron-Cohen et al., 1999; Happé, 1994). Further analysis would be done on faux pas-related questions that were divided into faux pas detection and faux pas understanding. Some researchers found that children developed faux pas recognition earlier than faux pas understanding. Jackson stated that late-development function disappeared earlier while early-development function was impaired later with aging or brain damage etc. (see Gulya et al., 2002). The hypothesis was that the performances on strange stories and faux pas detection questions in old adults were better than those on faux pas stories and faux pas understanding questions.

Finally, the correlations between two kinds of intelligence (crystallized and fluid intelligence) and theory of mind abilities, and the correlations between three kinds of IQs (Verbal, Performance and Full-scale IQs) and theory of mind abilities were computed. The Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Gong, 1992) was used to measure IQs of the participants and the WAIS scores are normally classified into the Verbal and Performance IQs. Also, crystallized and fluid intelligence were measured separately by Vocabulary Test and Digit Symbol Substitution Test of the WAIS-R (Maylor et al., 2002). Maylor et al. (2002) found that the worse performance of the old
group could not be fully accounted for by intelligence. However, Sullivan and Ruffman (2004) found that the performance in strange stories understanding was related to fluid intelligence, but the performance on video and emotion-labelling tasks was not. Hence, in the present study two different language-based theory of mind tasks were used to help clarifying the relationship between intelligence and theory of mind.

**Method**

**Participants**

There were 30 participants in the young as well as in the old group. All the old and young participants were recruited through two university databases and one graduate school database. The old participants were aged between 62 and 77 years (mean age = 70; 4 females and 26 males). The female sample of the old group was very small because few women could receive high education in 1930s and 1940s in China. The young participants were aged between 19 and 25 years (mean age = 22; 4 females and 26 males). The gender proportions of the two age groups were matched because females tend to perform better on theory of mind tasks (Bosacki, 2000; Charman, Ruffman, & Clements, 2002; Walker, 2005). All participants were healthy.

**Materials**

Three “strange stories” (theory of mind), three control stories, and three faux pas stories (theory of mind) were used. The “strange stories” were taken from Happé (1994), Happé et al. (1998), and Maylor et al. (2002) concerning double bluff, mistakes, and white lie. Concerning the control stories, one was taken from Happé et al. and the other two stories were newly constructed in a similar way. Concerning the faux pas stories, one was taken from Stone, Baron-Cohen, Calder, Keane, & Young (2003) and the other two stories were newly constructed in a similar way. These stories and the following questions were of similar length. The questions following the theory of mind stories required an inference about others’ thoughts and feelings. In contrast, the questions following the control stories did not require an inference about others’ mental state. Examples of the theory of mind and control stories are included in the Appendix A and B.

Nine stories and the questions following each strange story and control story were printed on separate pages, and were presented with a folder. Thus, participants were required to answer the questions by remembering the stories (e.g. “memory load condition” in Maylor et al., 2002). The order of the three types of stories was counterbalanced.

**Procedure**

The old participants were tested individually in their homes while the young participants were tested individually at university. All the stories were presented with a test booklet, with instructions written on the first page of the file.

Concerning the strange stories and the control stories, the participants were asked to read each story silently until they had understood them, turn to the next page for the test question, and answer the question. The participants’ answers were rated as follows: a score of 2 for a full and explicit correct answer, a score of 1 for a partial correct answer, and a score of 0 for a completely incorrect answer. A second marker who was blind to both the participants and hypothesis coded 30% of the answers. The first and second markers had a 94.4% agreement rate. Examples of scoring criteria for the stories are given in the Appendix A.

Concerning the faux pas stories, the participants were also asked to read each story silently until they had understood them, turn to the next page for a blank paper, and answer the four test questions (identical to Stone et al., 2003) and one control question the experimenter asked orally. One point was scored for the correct answer to the first question, another point for the correct answer to the second question. Participants proceeded to the control question immediately if they answered “no” to the first question. The control question was not scored and all the participants answered it correctly. The answers to the third and fourth questions were rated as follows: a score of 2 for the correct answer containing explicit mental state terms, a score of 1 for the correct answer containing explicit mental state with an additional prompt, a score of 0 for responding in error with an additional prompt. Examples of scoring criteria for the stories are given in the Appendix B.
RESULTS

Intelligence Measure

Table 1 shows the scores for crystallized and fluid intelligence, and the Verbal, Performance, and Full-scale IQs measured by the WAIS-R. Independent-samples $t$ tests revealed significant effects of age group for crystallized intelligence, $t(58) = 2.16, p < .05$, fluid intelligence, $t(58) = –10.45, p < .01$, and Performance IQ, $t(58) = 2.37, p < .05$, but not for Full-scale IQ, $t(58) = –0.09, ns$ and Verbal IQ, $t(58) = –1.70, ns$. Concerning the Vocabulary Test, the old group performed better than the young group. However, for Digit Symbol Substitution Test, the pattern was reverse, that is, the young group scored higher than the old group. Thus, in line with the findings of cognitive ability in the aged, the measures of crystallized and fluid intelligence showed the typical pattern of aging (Horn & Cattell, 1967), with a fall in fluid intelligence from youth to old age (Maylor et al., 2002; Sullivan & Ruffman, 2004), but a growth in crystallized intelligence from youth to young-old age (from 60 to 74 years) (Maylor et al., 2002). The Full-scale and Verbal IQs were matched. However, the Performance IQ of the old group was higher than that of the young group, which was inconsistent with the finding of Wechsler that old people scored lower than young people in Performance scale of the WAIS (see in Ye, 1991) perhaps because all the old participants in the present study were engineering professors and researchers.

Theory of Mind Tasks

The mean scores on strange, faux pas and control stories were summarized in Table 2. To compare scores on the three kinds of tasks, the scores of the strange and control stories (total score range: 0–6) were transformed to a comparable scale to that used for faux pas stories (total score range: 0–18) by multiplying the total scores of strange and control stories by 3 separately (Gopnik & Astington, 1988). ANOVA with age group
(young vs. old) and story type (control vs. strange vs. faux pas stories) as independent variables revealed a significant interaction between them, $F(2, 116) = 5.07, MSE = 10.68, p < .01$, and a main effect for story type, $F(2, 116) = 40.64, MSE = 10.68, p < .01$. Post hoc tests (Scheffé) showed that all the participants performed on the faux pas stories significantly worse than did on the strange and control stories ($p < .01$), while no significant difference existed in the performance on the strange and control stories. In addition, the main effect of age group was not significant, $F(1, 58) = 0.83, ns$. According to the interaction, further analyses were done on the faux pas, strange, and control stories separately. The old group performed worse than the young group on faux pas stories, $t(58) = –2.25, p < .05$, while the old group performed as well as the young group on strange stories, $t(58) = 1.46, ns.$ and on the control stories, $t(58) = 0.00, ns$.

The four faux pas-related questions following each story can be further broken down into two kinds of questions that determine whether or not the participant has correctly recognized the faux pas (“Did anyone say something awkward?” and “Who?”), and the follow-up questions that determine whether or not the participants understood the faux pas (“Why should they not have said it?” “Why do you think they did say it?”) (Stone et al., 2003). Thus, each participant can be given a “faux pas detection” score of 0–6 (for the old group, $M = 4.67, SD = 1.54$; for the young group, $M = 5.08, SD = 1.45$), and a “faux pas understanding” score of 0–12 (for the old group, $M = 6.17, SD = 3.50$; for the young group, $M = 8.35, SD = 2.86$). To compare scores on the two kinds of questions, the scores of the faux pas detection were transformed to a comparable scale to that used for the faux pas understanding by multiplying the total scores of the faux pas detection by 2 (Gopnik & Astington, 1988). ANOVA with age group (young vs. old) and question type (faux pas detection vs. faux pas understanding) as independent variables revealed a significant interaction between age group and question type, $F(1, 58) = 7.27, MSE = 2.12, p < .05$, and a significant main effect of the question type, $F(1, 58) = 84.90, MSE = 2.12, p < .01$. However, the main effect of the age group was not significant, $F(1, 58) = 3.70, ns$. Simple effects were calculated. The old group performed worse than the young group in faux pas understanding, $t(58) = –2.63, p < .05$, while the old group performed as well as the young group in faux pas detection, $t(58) = 0.95, ns$.

### Table 2. Mean Scores on Faux Pas, Strange and Control Stories in the Young and Old Group

<table>
<thead>
<tr>
<th></th>
<th>Young ($N = 30$)</th>
<th>Old ($N = 30$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strange stories (total score range: 0–6)</td>
<td>5.33 (1.21)</td>
<td>5.70 (0.65)</td>
</tr>
<tr>
<td>Control stories (total score range: 0–6)</td>
<td>5.67 (0.76)</td>
<td>5.67 (0.92)</td>
</tr>
<tr>
<td>Faux pas stories (total score range: 0–18)*</td>
<td>13.43 (4.16)</td>
<td>10.83 (4.77)</td>
</tr>
</tbody>
</table>

Note: Asterisks indicate significant differences between the two age groups, $* p < .05$. 

(232)
Correlation Analysis

The correlations were compared between different kinds of intelligence and performance on different stories, and between different IQs and performances on different stories (see in Table 3). The correlations between crystallized intelligence and the scores of strange stories were significant both for the old group, $r = 0.48$, $p < .01$, and for the young group, $r = 0.38$, $p < .05$. However, the correlation between crystallized intelligence and the performance on faux pas stories was only significant for the old group, $r = 0.52$, $p < .01$. In order to examine whether differences in crystallized intelligence could account for the differences in performance on faux pas tasks between the two age groups, a one-way analysis of covariance (with crystallized intelligence as a covariate) was computed. The results still indicated a significant main effect for age group, $F(1, 57) = 7.42$, $MSE = 19.19$, $p < .01$. Thus, crystallized intelligence could not account for the difference between the two age groups on faux pas stories. All the other correlations for the old and young group separately were not significant (all $ns$).

DISCUSSION

The results in the present study suggest that the abilities to understand strange stories and physical stories are preserved in old adults, but the old group showed worse performance on the faux pas stories, especially in the faux pas understanding. This suggests that the developmental trends from youth to old age were different on strange stories and faux pas stories.

On the strange and physical stories, the old group performed as well as the young group in the present study. The scores of both young and old adults on the two tasks approached the ceiling. It suggested that strange and physical stories were easy for the old and young people with higher education. This finding concerning the physical stories is in line with Happé et al. (1998), Maylor et al. (2002), and Sullivan and Ruffman (2004), who found no difference between the groups on physical stories. However, our findings concerning the strange stories are not in line with previous research which showed either a
fall (Maylor et al., 2002; Sullivan and Ruffman, 2004) or an increase (Happé et al., 1998) in performance on strange stories with growing age. This different finding could be due to the characteristics of the participants which might influence the developmental trend. The two age groups in the present study were well matched in education level and IQ, which was not the case in the above three studies. The background of the old group in our study was similar to that in Happé et al. All old participants in the present study were professors at a university and a graduate school with IQ ranging from 117 to 144, most of them still did research, while the old participants in Happé et al. were from the Research Center of the Department of Neurology, Boston University School of Medicine. We found that the performance on strange stories of the old group in the present and Happé et al. (mean score = 14.9, Full score = 16) study both approached the ceiling. In contrast, the background of the young participants in the two studies was different to some extent because the young adults in the present study were all university students with IQ ranging from 117 to 137.

On the faux pas stories, the result supported the conclusion of Maylor et al. (2002) and Sullivan and Ruffman (2004) that the old group performed worse than the young group. Stone et al. (1998) have developed a faux pas understanding task for adults which includes inferences about epistemic mental states, intentionality, and affective mental states simultaneously (Stone et al., 2003). The participant needs to understand that a story character has a false belief, and a person unintentionally says something awkward or hurtful to another person. From the present results, we found that faux pas stories are more difficult for old adults than strange and control stories, especially faux pas understanding. During the experiment, most old adults could recognize the faux pas, but some could not understand the faux pas. For example, they thought the character harmed in the story should not be sad because it was the other character’s fault. Moreover, they thought the character who harmed the other in the story did so intentionally. Perhaps their living experiences influenced their judgments. The finding indicated that children developed faux pas detection earlier than faux pas understanding (Wang & Su, under review). It seemed that the development and dissolution of faux pas detection and understanding could be to some extent described by the Jacksonian principle—a first-in, last-out principle.

The present results showed that the old group’s performance on strange stories and faux pas stories was not related to fluid ability, full-scale IQ, performance IQ, and verbal IQ, but to crystallized ability. Moreover, for the young group, only performance on strange stories correlated significantly with crystallized ability. The results of Experiment 2 in Maylor et al. (2002) also indicated that the correlation between theory of mind and crystallized ability was significant with age controlled. Sullivan and Ruffman (2004) expected that the performance of strange stories would correlate with crystallized ability in each age group because theory of mind tasks were mainly verbal tasks and the test of crystallized ability involves language ability. Many researchers found a strong link between theory of mind and language ability (Happé, 1995; Jenkins & Astington, 1996; Tager-Flusberg & Sullivan, 2000). Our results supported the expectation of Sullivan and Ruffman. However, crystallized ability could not account for the difference between the
old and young group on faux pas stories from a one-way analysis of covariance although there existed a high correlation between them. The performance on control stories was not related with crystallized ability in the old and young group separately although control stories were also language-based. We could not find a good explanation for this result. In addition, partially contrary to Sullivan and Ruffman, no correlations existed between fluid ability and performance on theory of mind tasks and control tasks. Moreover, full-scale, performance, and verbal IQ did not correlate with performances on faux pas, strange, and control stories for both old and young participants separately. This suggests that theory of mind performance can not be explained by fluid intelligence and three kinds of IQs. All these would lend supports to the argument of Happé et al. (1998) that a “dedicated processing mechanism for theory of mind” (p. 360) exists.

The limit of the present study is that the number of strange stories is small. Only three strange stories were used. However, Happé et al. (1998) used eight, Maylor et al. (2002) used five with memory loads, and Sullivan and Ruffman (2004) used six. Another limit is that an old-old age group should be added, like Maylor et al. (2002), to further explore the trend of theory of mind abilities.

In sum, the old adults in our study displayed different levels of theory of mind abilities on the Happé’s stories and faux pas stories. The performance on faux pas stories, especially in faux pas understanding, supports the findings that the old people performed worse than the young people in theory of mind abilities (Maylor et al., 2002; Sullivan & Ruffman, 2004). However, the old adults performed as well as the young adults on Happé’s stories. In addition, the performance on faux pas and strange stories for both the old and young adults was independent of fluid intelligence, full-scale IQ, verbal IQ, and performance IQ.

REFERENCES


Happé, F. 1995. The role of age and verbal ability in the theory of mind task performance of subjects with


(Manuscript received March 18, 2006; Revision accepted November 23, 2006)

Appendix A

Example strange story

The YELLOW army captured in a battle a tank commander from the BLUE army. They wanted to know from him where his troops were. They knew that the tanks were either in the military base or inside the cavern near that base. In addition to that, the YELLOW army believed that it was totally impossible to get the real location of his army from this officer due to his loyalty. In fact, this tank officer was indeed quite loyal to his army and he was quite clever as well. He was very sure that the enemy could never believe his “confession”. So he decided to tell the truth: “The tanks are inside the military base”.

Question: “Why did the officer tell the enemy the true location of the tanks?”

Example responses scored:

“Because he thought the enemy knew he would not tell them the truth.” (Score: 2)

“Because he would not tell them the truth.” (Score: 1)

Example control story

A thief was breaking into a jewelry shop. He skillfully unlocked the door and then carefully crept along the floor bypassing the scanning beam. If he touched any scanning beam, all the alarm bells in the jewelry shop would be triggered. Finally, the thief opened the coffer with the jewels shining inside. While he reached his
hands for these jewels, all of a sudden, the thief felt his feet stepped over something soft. He heard the sound “miaow” and something flocky all over the body ran along his side toward the door of the jewelry shop. And all the alarm bells rang.

Question: Why did the alarm bells ring?

Example responses scored:
“Because the burglar disturbed a cat, which ran through the scanning beam.” (Score: 2)
“Because something touches the scanning beam.” (Score: 1)

Appendix B

Example of faux pas story

Example: Qian bought her friend Ya a crystal bowl for a wedding gift. Ya had a big wedding and there were a lot of presents to keep track of. About a year later, Qian was over one night at Ya’s place for dinner. Qian dropped a wine bottle by accident on the crystal bowl, and the bowl shattered. “I’m really sorry, I’ve broken the bowl,” Qian said. “Don’t worry,” said Ya, “I never liked it anyway. Someone gave it to me for my wedding.”

Test questions:
The first question: “Did anyone say something he or she shouldn’t have said?” or “Did anyone say something awkward?”

Example responses scored:
Yes. (Score: 1)
No. (Score: 0)

The second question: “Who said something he or she should not have said?”

Example responses scored:
Ya (Score: 1)
Qian or Others (Score: 0)

The third question: “Why shouldn’t the individual in the story have said what he or she did?” (tests whether the participant understood that the listener would be hurt or insulted, an inference about affective mental states)

Example responses scored:
Because Qian is upset or sorrow. (Score: 2)
Answer correctly by an additional prompt, “What does Qian feel?” (Score: 1)

The fourth question: “Why do you think he or she did say it?” (tests whether the participant understood that the faux pas was unintentional, an inference about epistemic mental states and intentionality).

Example responses scored:
Because Ya does not know or realize she should not say it, or because Ya intends to console Qian. (Score: 2)
Answer correctly by an additional prompt, “Does Ya know or realize she should not say it?” or “What does Ya intend to do?”. (Score: 1)

Control questions: “What is the gift that Qian bought her friend Ya?”

Example correct responses (no scoring): crystal bowl