Psychology and Aging
2007, Vol. 22, No. 3, 639–643

0882-7974/07/$12.00 DOI: 10.1037/0882-7974.22.3.639

Exploring the Specificity of Age-Related Differences in Theory of Mind Tasks

Gillian Slessor, Louise H. Phillips, and Rebecca Bull
University of Aberdeen

Tasks assessing theory of mind (ToM) and non–mental state control tasks were administered to young and older adults to examine previous contradictory findings about age differences in mental state decoding. Age differences were found on a verbal ToM task after controlling for vocabulary levels. Older adults achieved significantly lower scores than did younger adults on static and dynamic visual ToM tasks, and a similar pattern was found on non-ToM control tasks. Rather than a specific ToM deficit, older adults exhibited a more general impairment in the ability to decode cues from verbal and visual information about people.

Keywords: social understanding, aging, theory of mind, emotion

Theory of mind (ToM) is the ability to represent mental states such as the beliefs, thoughts, and intentions of others (Maylor, Moulson, Muncer, & Taylor, 2002). Most previous studies assessing ToM have focused on developmental trends in young children or impairment in clinical populations with social deficits (e.g., in those with traumatic brain injury, schizophrenia, and autism; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Brüne, 2005; Channon & Crawford, 2000; Perner & Lang, 1999). Less is known about the changes in ToM that accompany adult aging. Age-related changes in neural functioning are particularly evident in the frontal and temporal lobes (Greenwood, 2000; West, 1996), the brain regions most often associated with ToM (Apperly, Samson, Chiavarino, & Humphreys, 2004; Frith & Frith, 2003), suggesting that ToM may decline with age. In contrast, it has been argued that older adults are well practiced in social skills such as representing the mental states of others (Happé, Winner, & Brownell, 1998) and may prioritize the processing of social and emotional information (Carstensen, Isaacowitz, & Charles, 1999).

In the first study to investigate age effects on ToM, Happé et al. (1998) administered a ToM stories task that required participants to make inferences about what a protagonist understood about the mental state of another individual by decoding subtle cues to sarcasm, deception, and social rule violation. Performance of older adults was significantly better than that of younger participants, whereas no age difference was found on a control stories task containing no ToM component. According to Happé et al. (1998), wisdom and social intelligence increase with age, resulting in superior ToM. A further study found no age difference in the ability to distinguish from stories whether someone had committed a social faux pas (MacPherson, Phillips, & Della Sala, 2002).

However, recent studies using similar verbal tasks have found that ToM declines with age. Maylor et al. (2002, Experiment 1) compared performance of three age groups (with mean ages of 19, 67, and 81 years) on the ToM stories task. ToM in both older groups was impaired, but when memory demands were alleviated, ToM impairment disappeared for the young-old group. Sullivan and Ruffman (2004) replicated these age declines in the ToM stories task and also found that older adults performed poorly on a ToM videos task that assessed the ability to decode thoughts and feelings from short, dynamic video clips. Phillips, McLean, and Allen (2002) found that older adults performed worse than young adults on the ToM eyes task (Baron-Cohen et al., 2001), in which mental state judgments must be made about pictures of the eye region. Saltzman, Strauss, Hunter, and Archibald (2000) investigated the effect of aging on a range of ToM tasks, including stories and cartoons tasks, and found that older participants had a lower composite ToM score.

Age-related declines in ToM were also found when stories were used to assess complex second-order ToM, requiring participants to consider the thoughts of two different characters, although first-order ToM, which involved consideration of only one character's perspective, was not impaired with age (McKinnon & Moscovitch, 2007). This could indicate that rather than showing a specific deficit in the ability to represent mental states, older adults show a decline in the domain-general resources required for more complex ToM judgments. German and Hehman (2006) also found that age declines were greatest on a *false belief* version of a ToM stories task where the protagonist in a story had a mistaken understanding about the state of the world.

Therefore, previous studies examining age-related changes in ToM have produced a contradictory pattern of results. However, a number of methodological issues exist with these studies. The ToM tasks used in these studies have often been subject to ceiling effects (e.g., German & Hehman, 2006; Keightley, Winocur, Burianova, Hongwanishkul, & Grady, 2006; MacPherson et al., 2002; McKinnon & Moscovitch, 2007; Saltzman et al., 2000). Also,

Gillian Slessor, Louise H. Phillips, and Rebecca Bull, School of Psychology, University of Aberdeen.

This research was supported by scholarships from the Carnegie Trust and Wellcome Trust.

Correspondence concerning this article should be addressed to Gillian Slessor, School of Psychology, College of Life Sciences and Medicine, Aberdeen University, Aberdeen, AB24 2UB, Scotland, United Kingdom. E-mail: gillian.slessor@abdn.ac.uk

some studies have had small samples, raising questions about the reliability of the findings (e.g., the number of older participants ranges from 8 to 20 in German & Hehman, 2006; Happé et al., 1998; McKinnon & Moscovitch, 2007; and Saltzman et al., 2000). The small number of items used in some studies (e.g., German & Hehman, 2006; Keightley et al., 2006; Maylor et al., 2002) is also cause for concern.

An important theoretical issue in measuring ToM is the use of appropriate control tasks to distinguish whether age effects reflect specific failures to decode mental state or more general changes in understanding complex texts and perceiving visual information. This issue has previously been addressed in the ToM stories task. Despite evidence of age-related improvements in performance on a ToM stories task, Happé et al. (1998) found no age deficit on a control stories task involving no ToM component. Conversely, using the same stimuli, Maylor et al. (2002) and Sullivan and Ruffman (2004) both found evidence of greater age-related decline in performance on ToM stories tasks compared with control physical stories tasks. German and Hehman (2006) constructed vignettes where participants were asked to make judgments about either the social agent (ToM task) or the mechanical device (control task) featured in the story. Older adults were impaired on both the ToM and the control tasks, providing evidence against a specific age-related decline in ToM (see also Keightley et al., 2006). Studies of age differences in visual tasks of ToM (Phillips et al., 2002; Sullivan & Ruffman, 2004) have not used non-ToM control conditions to investigate whether the age changes are specific to ToM decisions.

The majority of aging studies of ToM have been limited by only considering one aspect of ToM. Saltzman et al. (2000) argued that different ToM measures assess different aspects of social understanding, which might be differentially affected by age, making it essential to assess performance on a range of ToM tasks in the same sample. It seems particularly important to gain more information about the effects of age on analyzing dynamic social stimuli, such as Sullivan and Ruffman's (2004) videos task, as this is more akin to real-life situations and interactions.

The present study therefore addresses the following research questions: First, does aging cause declines in all aspects of ToM judgment? Previous studies have indicated a mixed pattern of age effects on ToM stories tasks and declines on visual ToM tasks, but none has compared aging effects on verbal, static visual, and dynamic visual tasks of ToM within the same sample. In the present study, three ToM tasks are used. The stories and eyes tasks are the most commonly used ToM tasks in adult research, and populations with known deficits in ToM (e.g., with autism and Asperger's syndrome) are severely impaired on these tasks (Baron-Cohen et al., 2001; Heavey, Phillips, Baron-Cohen, & Rutter, 2000). The videos task has only been used in one previous aging study (Sullivan & Ruffman, 2004); however, it is similar in format to the eyes task, suggesting that it will successfully assess ToM ability. Second, is there evidence for age impairment specific to ToM performance? To assess this, for each ToM task, we included matched control tasks that did not require ToM judgments. To deal with problematic issues in previous studies of age and ToM, we had a larger sample size in the current study, piloted tasks to avoid ceiling effects, and used at least 10 items per task.

Method

Participants

Two groups of participants were recruited: 40 young adults (28 female) ranging in age from 16 to 40 years ($M = 20.08$ years, $SD = 4.64$), the majority being students who completed the study for course credit, and 40 older adults (25 female) ranging in age from 60 to 74 years ($M = 66.95$ years, $SD = 4.31$), recruited through the local participant panel and reimbursed £8 (approximately $16) for their time. All had good command of English and were free of past or present neuropsychological disorders. The groups did not differ in the number of years of education, $t(78) = -0.257$ (young adult $M = 13.66$ years, $SD = 1.68$; older adult $M = 13.51$ years, $SD = 3.28$). All participants completed the first 30 items, each worth 1 point, from the Mill Hill Vocabulary Scale (Raven, Raven, & Court, 1998), with older adults achieving a higher vocabulary score, $t(78) = 9.21$, $p < .001$ (young adult $M = 15.83$, $SD = 2.8$; older adult $M = 22.48$, $SD = 3.58$).

Materials and Procedure

Six computerized tasks were administered, three ToM and three control. The order in which the different ToM and control tasks were presented was counterbalanced across participants. In addition to verbal and written instructions, one practice item preceded each task to familiarize participants with task demands. For all tasks, participants responded using labeled keys on the keyboard.

Stories tasks. A modified multiple-choice stories task was used. There were 12 ToM stories, which concerned double bluffs, mistakes, and white lies (based on stimuli from Channon & Crawford, 2000; Happé et al., 1998; and Stone, Baron-Cohen, & Knight, 1998). Each story was followed by a question regarding the intentions of the person in the story and four possible response options labeled *A–D* (see Bull, Phillips, & Conway, 2007, for details of the construction of this task). Twelve control stories were also administered that had similar properties but did not contain a ToM component, instead requiring general inferences, for example, about physical or mechanical causation (based on stimuli from Happé et al., 1998, and Stone et al., 1998). To reduce memory load, we presented the passage, question, and response options on-screen at the same time, where they remained until the participant had responded.

Videos tasks. The ToM videos task consisted of 16 silent 5-s color video clips adapted from Sullivan and Ruffman (2004) that portrayed characters interacting. Participants were instructed to choose the word that best described the thoughts or feelings of the person in the video. It was clear from the angle of filming which character participants were to judge. Each clip was surrounded by four possible options describing mental state such as *frustrated*, *excited*, *annoyed*, and *bored*. Incorrect response options were generated from free responses made by young adults viewing the video clips in a pilot study. A control videos task using the same stimuli was also administered where participants chose the option that best described the age and gender of the key character (e.g., *Male 40–50*, *Female 40–50*, *Female 50–60*, *Male 50–60*). In both tasks, the options appeared before, during, and after the video clip.

Eyes tasks. The ToM eyes task (Baron-Cohen et al., 2001) consisted of 25 black-and-white photographs of the eye region.

These were surrounded by four words, one target and three foils, describing complex mental states such as *annoyed*, *hostile*, *horrified*, and *preoccupied*. Participants were asked to identify which word best described the mental state of the person. The control eyes task used the same stimuli but required participants to judge the age and gender of the target person (response choices were similar to those used in the control videos task).

Results

Percentage accuracy and standard deviations for correct responses on each task (ToM and control) for each age group (young and old) are shown in Table 1. A series of separate mixed-design analyses of variance (ANOVAs) with two levels of task (ToM vs. control) as the within-subjects variable and age category as the between-subjects variable were conducted individually for the stories, eyes, and videos tasks. As the two age groups differed substantially in vocabulary levels, for each task, an analysis of covariance (ANCOVA) was carried out with vocabulary score as a covariate (adjusted means following ANCOVA are also shown in Table 1).

For the stories task, the ANOVA revealed no significant effect of age, $F(1, 78) = 2.72$, $p = .10$, $\eta_p^2 = .03$, or task (ToM vs. control), $F(1, 78) = 1.42$, $p = .288$, $\eta_p^2 = .01$, and no significant interaction between age and task $F(1, 78) < 1$, $\eta_p^2 = .00$. Similar performance patterns were found for the ToM and control tasks with no evidence of any age-related decline. However, ANCOVAs controlling for age differences in vocabulary revealed an age-related decline in performance on the ToM stories task, $F(1, 77) = 11.24$, $p < .01$, and a nonsignificant decline on the control stories task, $F(1, 77) = 3.26$, $p = .08$.

Analysis of the videos task revealed a significant main effect of age, $F(1, 78) = 19.96$, $p < .001$, $\eta_p^2 = .20$. There was no Task × Age interaction, $F(1, 78) < 1$, $\eta_p^2 = .00$, and no main effect of task, $F(1, 78) = 1.60$, $p = .210$, $\eta_p^2 = .02$. On the eyes task, there were significant main effects of age, $F(1, 78) = 13.61$, $p < .001$, $\eta_p^2 = .15$, and task, $F(1, 78) = 46.91$, $p < .001$, $\eta_p^2 = .38$, with accuracy higher in the ToM compared with the control condition, but no significant interaction between task and age, $F(1, 78) = 1.49$, $p = .226$, $\eta_p^2 = .02$. The results show that older participants performed more poorly than young participants on the ToM videos and eyes tasks and their respective control tasks. However, the lack of significant interactions suggests that they were not disproportionately impaired on the ToM tasks in comparison to the control tasks. For both the eyes and the videos tasks, ANCOVAs including vocabulary score as a covariate resulted in no changes to the overall pattern of age effects (see Table 1 for adjusted means).

Discussion

Our aim in this study was to investigate whether there was evidence for a specific age-related decline in ToM and, if so, whether aging results in decline in all aspects of ToM judgments, including those involving verbal material (stories task) and static (eyes task) and dynamic (videos task) visual stimuli. For the stories task, no main effect of age was found and there was no interaction with task type (ToM vs. control). These findings suggest that older adults have no problems decoding information from verbal material, even when it involves a ToM judgment. This is consistent with previous research that found no evidence of age-related changes on variants of the ToM stories task (MacPherson et al., 2002; Phillips et al., 2002; Saltzman et al., 2000). In addition, the results of the current study are also consistent with those of Maylor et al. (2002, Study 1), who found no age differences in performance on a ToM stories task when the older participants were of an age range similar to that in the current study and memory demands of the task were minimized. However, these results are in direct contrast to the findings of Keightley et al. (2006) and German and Hehman (2006), who found that aging resulted in an overall decline in the ability to process information from verbal material, as their older groups of participants demonstrated poorer performance on both the ToM true belief tasks and the control stories tasks used in these studies. In addition, the findings of the present study are also inconsistent with previous suggestions of age-related improvements in performance on the ToM stories task (Happé et al., 1998) or the finding of an age-related decline in performance on this ToM measure (McKinnon & Moscovitch, 2007; Sullivan & Ruffman, 2004). A puzzle remains regarding the effects of aging on the ToM stories task, and one possible explanation lies in the varying vocabulary levels of the participants in different studies.

In the present study, the older participants had significantly superior vocabulary scores in comparison to the younger group, which could have accounted for the lack of age-related decline in the stories task. Statistical analysis with vocabulary as a covariate supported this, revealing a significant age-related deficit on the ToM stories task and an age effect on the control stories that approached significance. Therefore, in the current sample, it seems likely that older adults' advantage in vocabulary may have counteracted age declines in extracting information in verbal tasks, in

Table 1
Accuracy (Mean % Correct) and Standard Deviations for Theory of Mind (ToM) and Control Tasks and Adjusted Means Controlling for Mill Hill Vocabulary Scores

	Young			Older		
Task	% correct	Adjusted % correct	*SD*	% correct	Adjusted % correct	*SD*
ToM stories	78.13	83.58	13.03	75.21	69.73	14.56
Control stories	80.63	81.85	9.88	76.67	75.44	11.97
ToM videos	64.06	65.14	11.21	56.10	55.01	11.28
Control videos	67.19	67.51	11.65	57.50	57.18	13.44
ToM eyes	72.30	75.32	10.58	66.70	63.68	11.15
Control eyes	61.70	61.24	14.20	51.50	51.95	14.58

particular when they were required to make mental state judgments. These results suggest that there may be a more general rather than ToM-specific age-related decline in the ability to make judgments from verbal material, which concurs with findings of age deficits on both ToM true belief tasks and control stories tasks (German & Hehman, 2006; Keightley et al., 2006).

Inspection of the population samples of past research reveals an interesting pattern with regard to the relation between age-related changes on stories task performance and vocabulary level. For example, in three previous studies that found age-related performance deficits on ToM stories tasks (German & Hehman, 2006; Maylor et al., 2002; Sullivan & Ruffman, 2004), there was no significant difference between the vocabulary scores of young and old participants. Details of the vocabulary level of participants were not provided by the other cited studies. However, Happé et al. (1998), who found age-related improvements in ToM, described their old sample as being of "higher general intelligence" (p. 360), which suggests that the vocabulary of their older participants may have been superior to that of the younger participants. Such diversity in the vocabulary levels of samples in different studies may account for the contradictory findings in the literature concerning age differences in performance on the ToM stories task. This highlights the importance of obtaining background information from samples in aging studies of social cognitive abilities.

These influences of vocabulary were not found on two visual tasks assessing ToM, the eyes and videos tasks, despite the complex terms used to describe mental state in some items. The older participants performed significantly worse than the younger participants did on both of these tasks, suggesting that older participants may have an impairment in visual ToM tasks that require the ability to attribute thoughts and feelings to another person (for similar findings, see Phillips et al., 2002; Sullivan & Ruffman, 2004). Conversely, these results are not in keeping with McKinnon and Moscovitch (2007), who suggested that basic, first-order ToM processes, which are assessed in the eyes and videos tasks, remain intact with aging. However, McKinnon and Moscovitch (2007) used a different task than those used in the present study (i.e., the task used by McKinnon & Moscovitch, 2007, required participants to judge one character's perspective from verbal material, whereas the tasks used in the current study assessed the ability to decode the thoughts and feelings of others from visual stimuli). Moreover, the first-order ToM task used by McKinnon and Moscovitch (2007) was subject to ceiling effects, making it difficult to interpret the lack of an age effect. The current findings also provide clear evidence that, similar to the age differences in the stories task, age-related deficits on visual ToM tasks are no greater than age declines on matched control tasks that do not require mental state judgments. These findings concur with McKinnon and Moscovitch (2007), who proposed that ToM was not a modular construct but also drew on general cognitive skills such as working memory capacity, switching between mental sets, and inhibition of prepotent responses. It is possible that the current pattern of age-related impairments on ToM and control tasks is due to problems in inhibiting irrelevant information or maintaining information in working memory (De Beni & Palladino, 2004; Reimers & Maylor, 2005; Sweeney, Rosano, Berman, & Luna, 2001). Such general impairments of cognition would affect extraction of information and subsequent decision making, leading to poorer performance on both the ToM and the control tasks.

However, it is also possible that the pattern of results found here reflects age differences in making social judgments. It is important to note that although the control eyes and videos tasks did not require mental state understanding, the judgments did require social understanding. For example, the judgment of age can be very important for social interaction (i.e., in the way in which people choose to address others, select conversation topics, and portray emotions). This explanation is in keeping with the finding of an age-related deficit in performance that approached significance on the control stories task, as the ability to abstract socially relevant information was not required to the same extent in this task. Future researchers should focus on developing appropriate ToM control tasks that do not require the extraction of information important to social understanding or the use of social stimuli. This would ascertain whether the age-related decline in these tasks is due to a general impairment in cognition or whether the deficit is specific to the abstraction of only social information.

The tasks used in the current study are the most common measures of ToM in adult populations. However, because all of the tasks involve decisions about a broad range of ToM aspects, including mental states, emotions, logical thought, and intentions, it would be useful in future research to develop tasks that assess more specific aspects of ToM and do not rely heavily on other cognitive abilities found to decline with age. The challenge is to develop such tasks while avoiding ceiling effects in adult performance.

In summary, an age-related deficit was found on both the ToM and the control versions of the eyes and videos tasks but not on the stories task, suggesting that older adults are worse than younger adults at decoding social information from static and dynamic visual displays but not from text. However, when vocabulary scores were covaried, a significant age-related decline in performance on ToM stories tasks and a deficit in performance on control stories tasks that approached significance were revealed, suggesting that differences in vocabulary may mask age-related differences on verbal tasks. Questions remain as to whether age-related declines in these tasks reflect general cognitive decline or more specific deficits in the ability to interpret and process social information and stimuli.

References

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16,* 1773–1784.

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 and high-functioning autism. *Journal of Child Psychology and Psychiatry, 42,* 241–251.

Brüne, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulletin, 31,* 21–42.

Bull, R., Phillips, L. H., & Conway, C. (2007). *The role of control functions in mentalizing: Dual task studies of theory of mind and executive functioning*. Manuscript submitted for publication.

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist, 54,* 165–181.

Channon, S., & Crawford, S. (2000). The effects of anterior lesions on

performance on a story comprehension test: Left anterior impairment on a theory of mind–type task. *Neuropsychologia, 38,* 1006–1017.

De Beni, R., & Palladino, P. (2004). Decline in working memory updating through aging: Intrusion error analyses. *Memory, 12,* 75–89.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B, 358,* 459–473.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101,* 129–152.

Greenwood, P. M. (2000). The frontal aging hypothesis evaluated. *Journal of the International Neuropsychological Society, 6,* 705–726.

Happé, F. G. E., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology, 34,* 358–362.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The awkward moments test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders, 30,* 225–236.

Keightley, M. L., Winocur, G., Burianova, H., Hongwanishkul, D., & Grady, C. L. (2006). Age effects on social cognition: Faces tell a different story. *Psychology and Aging, 21,* 558–572.

MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function, and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging, 17,* 598–609.

Maylor, E. A., Moulson, J. M., Muncer, A. M., & Taylor, L. A. (2002). Does the performance on theory of mind tasks decline in old age? *British Journal of Psychology, 93,* 465–485.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102,* 179–218.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3,* 337–344.

Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive approaches. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 57,* P526–P530.

Raven, J., Raven, J. C., & Court, J. H. (1998). *The Mill Hill Vocabulary Scale, Form 1 Senior*. Oxford, England: Oxford Psychologists Press.

Reimers, S., & Maylor, E. A. (2005). Task switching across the life span: Effects of age on general and specific switch costs. *Developmental Psychology, 41,* 661–671.

Saltzman, J., Strauss, E., Hunter, M., & Archibald, S. (2000). Theory of mind and executive functions in normal human aging and Parkinson's disease. *Journal of the International Neuropsychological Society, 6,* 781–788.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640–656.

Sullivan, S., & Ruffman, T. (2004). Social understanding: How does it fare with advancing years? *British Journal of Psychology, 95,* 1–18.

Sweeney, J. A., Rosano, C., Berman, R. A., & Luna, B. (2001). Inhibitory control of attention declines more than working memory during normal aging. *Neurobiology of Aging, 22,* 39–47.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292.

Received October 27, 2006
Revision received April 20, 2007
Accepted April 23, 2007 ■