Experimental Aging Research, 37: 481–502, 2011

ISSN: 0361-073X print/1096-4657 online
DOI: 10.1080/0361073X.2011.619466

THEORY OF MIND THROUGH THE AGES: OLDER AND MIDDLE-AGED ADULTS EXHIBIT MORE ERRORS THAN DO YOUNGER ADULTS ON A CONTINUOUS FALSE BELIEF TASK

Daniel M. Bernstein

Department of Psychology, Kwantlen Polytechnic University, Surrey, British Columbia, Canada

Wendy Loken Thornton

Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada

Jessica A. Sommerville

Department of Psychology, University of Washington, Seattle, Washington, USA

Theory of mind (ToM), or the ability to understand mental states, is a fundamental aspect of social cognition. Previous research has documented marked advances in ToM in preschoolers, and declines in ToM in older-aged adults. In the present study, younger (n = *37), middle-aged* (n = *20), and older* (n = *37) adults completed a continuous false belief task measuring ToM. Middle-aged and older adults exhibited more false*

Received 3 August 2009; accepted 26 July 2010.

This work was supported by a grant from the Social Sciences and Humanities Research Council of Canada to D.M.B. The authors would like to thank Reema Jayaker for assistance with data collection.

Address correspondence to Daniel M. Bernstein, Department of Psychology, Kwantlen Polytechnic University, 12666 72nd Avenue, Surrey, BC V3W 2M8, Canada. E-mail: Daniel.Bernstein@kwantlen.ca

belief bias than did younger adults, irrespective of language ability, executive function, processing speed, and memory. The authors conclude that ToM declines from younger to older adulthood, independent of age-related changes to domain-general cognitive functioning.

Theory of mind (ToM) refers to the ability to understand a variety of different mental states, such as beliefs, desires, intentions, emotions, and perceptions. ToM is a fundamental aspect of social cognition that guides social interactions and individual behavior across a range of contexts. There are marked improvements in children's ToM abilities during the preschool years, and in particular in children's ability to appreciate that they and others can hold beliefs about the world that are false (Wellman, 1990). Specifically, whereas 3- and 4-year-old children appear to think that their own and others' beliefs about the world directly reflect reality, 5-year-old children appreciate that beliefs are reflections of reality and therefore can be accurate or inaccurate.

Although there is a large literature on children's ToM, and on children's appreciation of false belief in particular, considerably fewer studies have examined ToM in older children, adults, and the elderly. A variety of evidence points to potential limitations in adult's ToM, particularly in situations in which adults must ignore their own knowledge, in order to consider the mental states of another, naïve individual (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Keysar, Lin, & Barr, 2003; Wilson & Brekke, 1994). Indeed, despite the marked shifts that are present in children's mental state reasoning during the preschool period, the life span developmental literature has shown that ToM continues to develop in a gradual fashion beyond the preschool years into young adulthood (Bernstein, Atance, Loftus, & Meltzoff, 2004; Birch & Bloom, 2007; Carpendale & Chandler, 1996; Maylor, Moulson, Muncer, & Taylor, 2002; Slessor, Phillips, & Bull, 2007; Wellman, Cross, & Watson, 2001). What remains contentious, however, is how and when mental state reasoning changes across adulthood.

Changes to Mental State Reasoning in Adulthood

Currently, there is mixed evidence regarding age-related changes to ToM abilities from younger to older adulthood, perhaps due to variability in the types of task and aging samples used across experiments. Whereas some studies suggest that older adults outperform younger adults (Happe, Winner, & Brownell, 1998), other work has yielded either no difference between younger and older adults (Keightley, Winocur, Burianova, Hongwanishkul, & Grady, 2006; MacPherson, Phillips, & Della Sala, 2002) or evidence that younger adults outperform older adults (Maylor et al., 2002). On balance, the literature to date shows that older adults have more trouble reasoning according to others' mental states than do younger adults (Slessor et al., 2007). Assuming that age-related declines in ToM are present, we might ask when age-related declines in ToM first emerge. Although most researchers studying mental state reasoning in aging populations include an older age group starting at age 60, relatively few studies have assessed changes to ToM in middle age (but see Charlton, Barrick, Markus, & Morris, 2009; Pardini & Nichelli, 2009).

Charlton et al. (2009) tested changes to ToM from middle age to older adulthood in a group of adults 50 to 90 years of age and found that ToM ability declined with age; however, their omission of younger adults prevents us from determining when ToM ability begins to decline in adulthood. To answer this question, Pardini and Nichelli (2009) tested differences in ToM from younger to older adulthood in four age groups (20–25; 45–55; 55–65; 70–75), and found that "decline in ToM capacities becomes statistically relevant after 55 years" (p. 103). These findings point to advanced middle age as the time point during which mental state reasoning ability begins to decline.

However, Pardini and Nichelli's (2009) study was limited in several ways. First, participants completed only one measure, the Revised Mind in the Eyes Test (translated into Italian and then validated). This task requires participants to judge the complex emotion that a person is exhibiting (for example, suspicion), based solely on a cropped photograph of that person's eyes. Participants indicate their answers by choosing one of four emotion words appearing with the photograph. The ability to identify others' emotional states based on limited evidence is an important aspect of ToM. However, other integral aspects of ToM, such as the ability to reason according to another perspective when that perspective conflicts with one's own, were not assessed. It is possible that different aspects of mental state reasoning follow different developmental trajectories. Indeed, some of these abilities may be spared in aging populations, whereas others may not. Furthermore, the authors did not administer tasks known to correlate with ToM, such as those that measure executive function and language ability. Indeed, many of the age differences in ToM reported in the literature disappear after controlling for executive function, in particular, processing speed, inhibitory control, and working memory (Bailey & Henry, 2008; German & Hehman, 2006; Maylor et al., 2002;

McKinnon & Moscovitch, 2007; Sullivan & Ruffman, 2004). Thus, it is impossible to know the extent to which age-related differences in ToM reported in the Pardini and Nichelli (2009) study reflect declines in domain-general cognitive abilities such as executive function versus differences in ToM per se.

For example, McKinnon and Moscovitch (2007) showed that older adults perform more poorly than younger adults on second-order false belief tasks (in which one maintains in mind and compares divergent beliefs or perspectives of different people, e.g., "Sally thinks that John thinks that the ball is under the blanket"). No age differences emerged on first-order false beliefs (in which one considers the thoughts and feelings of one person only). The authors interpret this data pattern as support for the idea that older adults perform more poorly than young adults on ToM tasks with high central processing demands, especially working memory. To test this claim, the authors conducted several follow-up experiments in which they manipulated working memory in a dual task involving younger adults performing false belief tasks. These experiments showed clearly that working memory load decreased false belief performance in young adults. From this, the authors concluded that working memory accounts for age differences in ToM between younger and older adults.

Although this study offers evidence for the role of working memory in ToM in adults, the evidence is indirect. Working memory was manipulated in younger adults only. The authors' argument for the mediating role of working memory in age-related differences in ToM rests upon the resemblance between the decline in false belief performance in younger adults under divided attention and the decline in performance in older adults under conditions of full attention.

More direct evidence for the mediating role of executive function in age-related declines in ToM comes from German and Hehman (2006). These authors administered a series of ToM tasks to younger and older adults. The ToM tasks involved stories that varied in terms of whether a protagonist acted on a true or a false belief (mental state). The authors also manipulated executive selection demands by changing whether the protagonist acted on the true or false belief by approaching or avoiding an object. Participants also completed a battery of executive function tasks. The authors found age-related declines in performance on the stories involving false belief, and these differences correlated most highly with measures of processing speed and inhibition (and to a lesser extent, working memory). The authors concluded that age-related declines in ToM are due in part to age-related differences in executive function. However, unlike McKinnon and Moscovitch (2007; see also Bailey & Henry, 2008), the authors did not report whether the age-related differences in ToM persist after controlling for executive function.

These limitations motivated us to investigate a central aspect of mental state reasoning, the ability to reason from another person's inaccurate or false belief, in adults at a range of different ages. We also administered a battery of cognitive tasks to determine whether age differences in false belief understanding reflect domain-general cognitive processes. Although several tasks have been used to assess ToM in aging populations (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Happe et al., 1998), these tasks are limited in their ability to investigate mental state reasoning across the life span. First, these tasks often show inconsistent performance when comparing younger to older adults, perhaps due to language and task demands, as well as ceiling effects (Slessor et al., 2007). Second, these tasks cannot be used with young children, because of the sophisticated language and cognitive demands of the task. Thus, existing methods limit the ability to draw conclusions about the development of ToM across the life span.

Reconsidering Tasks Used to Assess False Belief

An influential perspective in the ToM literature argues that changes to children's performance on false belief tasks between 3 and 5 years of age reflect qualitative conceptual change in children's understanding of beliefs (e.g., Wellman et al., 2001). The majority of studies that have been used to test false belief understanding in children require a discrete, categorical response. For example, in a classic change of location task, children are required to indicate which of two hiding locations a protagonist will search at after an object has been moved from one location to another without the protagonist's knowledge. Before roughly 4.5 years of age, children typically indicate that the protagonist will search where the object actually is. At 4.5 years of age, and later, children correctly indicate that the protagonist will search for the object where he or she believes it to be. The traditional interpretation of children's performance on this task is that whereas young children conceive of beliefs as mirrors of reality, older children understand beliefs as representations of reality and consequently understand that beliefs can sometimes be inaccurate or incorrect.

Another possibility, however, is that changes in terms of children's ability to represent and reason happen in a graded, continuous fashion, rather than in a discrete, either-or fashion. The discrete, categorical responses required by traditional false belief tasks may fail to capture developmental improvements or declines adequately (Bernstein, Atance, Meltzoff, & Loftus, 2007), if the underlying ability undergoes a subtle quantitative shift rather than a more dramatic qualitative shift.

Heeding these limitations, researchers have developed a task to measure mental state reasoning, and in particular false belief reasoning, on a continuum throughout the life span (Sommerville, Bernstein, & Meltzoff, 2010). The task was based on the well-known change-of-location task used with preschoolers (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). In the standard change-of-location task, a character (Sally) places an object (ball) in one location (box) and then leaves the room. While Sally is gone, another character (Anne) removes the object and places it in another location (cupboard). When Sally returns, children are asked a critical question that taps ToM, namely, "Where will Sally look for the ball?" Children are also asked control questions, such as "Where is the ball really?" or "Where did Sally put the ball before she left?" This task measures children's ToM understanding in discrete terms: One either answers correctly that Sally will look for the ball where she placed it, namely in the box, or one answers incorrectly that Sally will look for the ball where it is now, namely in the cupboard. In the continuous version of this task, the same scenario is enacted; however, this time the object is placed in a large (5-foot-long) box filled with Styrofoam peanuts. Rather than discrete locations like boxes and cupboards, the object is placed in different locations in one large container. Thus, this task provides a sensitive measure of the extent to which participants can ignore or disregard their own beliefs in order to reason from an ignorant perspective.

In the first study to use this continuous false belief task, preschoolers and young adults made more errors, and thus showed more bias, when the critical question involved a false belief ("Where will Sally look for the ball?") than in various control conditions (e.g., "Where did Sally put the ball before she left?"). Moreover, as could be expected from the development of mental state reasoning, adults showed less false belief bias than did 5-year-olds who, in turn, showed less bias than 3-year-olds. Also, preschoolers' false belief bias correlated significantly and positively ($r_{pb} = .46$) with their performance on the standard change of location task (Sommerville et al., 2010). The magnitude of this correlation is similar to that found between standard false belief tasks in children (see Carlson & Moses, 2001). Thus, the continuous false belief task appears to be sensitive to subtle developments in mental state reasoning, and is a good candidate for investigating mental state reasoning across a broad range of ages.

In the present work, we administered the continuous false belief task (Sommerville et al., 2010) to younger, middle-aged, and older adults. In addition, we included several measures that have been shown to correlate with ToM in children and adults, most notably tasks that tap language ability and executive function (Bull, Phillips, & Conway, 2008; Carlson & Moses, 2001; German & Hehman, 2006; Milligan, Astington, & Dack, 2007). Examining whether mental state reasoning is uniquely impaired in the elderly, or whether such impairment relates more broadly to other domain-general cognitive abilities such as language and executive function, can shed light on the nature of mental state reasoning in aging populations.

Based on prior work using categorical tasks in which declines in mental state reasoning emerge in the fifth decade (Pardini & Nichelli, 2009), we predicted that, in comparison to younger adults, middle-aged and older adults would show greater false belief bias on a continuous false belief task (hereafter called the Sandbox task). Moreover, to the extent that middle-aged and older participants do show false belief deficits, it is unclear whether these deficits are unique to false belief and mental state reasoning specifically, or whether they can be accounted for by declines in more domain-general cognitive abilities such as executive function and memory. Thus, our second research question was whether age-related declines in mental state reasoning would be related to their performance on other cognitive tasks. Based on inconsistent evidence in the literature linking deficits in mental state reasoning to domain-general cognitive abilities in aging populations (Charlton et al., 2009; Sullivan & Ruffman, 2004), we made no firm prediction about the link between false belief performance on the Sandbox task and traditional cognitive measures (executive function, processing speed, memory, and language ability). If mental state reasoning deficits are independent of domain-general cognitive abilities, then false belief performance on the Sandbox task should be unrelated to performance on traditional cognitive measures. However, if mental state reasoning deficits are due to domain-general cognitive abilities, then false belief performance on the Sandbox task should be related to individual differences on cognitive tasks known to decline with increasing age (Craik & Bialystok, 2006).

METHODS

Participants

Thirty-seven older adults (76% female; 89% right-handed; mean age = 67.6, *SD* = 6.0, range = 60–85) and 20 middle-age adults (80% female; 95% right-handed; mean age = 56.3, *SD* = 2.3, range = 51–59) from the greater Vancouver area were recruited through advertisements posted at local recreational centers and through presentations given at various community centers throughout the lower mainland of British Columbia. Thirty-eight younger adult participants (76% female; 97% right-handed; mean age = 19.2, $SD = 1.42$, range = 17–22) were recruited through an undergraduate university subject pool. Participants were ineligible if they had a history of major illnesses with known direct central nervous system effects (e.g., stroke, head injury, central nervous system malignancies, Parkinson's disease), or had previously identified cognitive impairments (e.g., diagnosis of dementia). All participants obtained scores of 25 or greater on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975), exceeding the cutoff commonly used for screening dementia (Kukull et al., 1994). All participants scored below the cutoff of mild depression on the Center for Epidemiologic Studies of Depression Scale (CES-D).

Additional exclusion criteria included a current diagnosis or history of a major psychiatric illness. Due to the visual nature of some tasks, we screened participants' visual acuity, with the lower limit of correction set at 20/50. All participants were tested individually in the same University laboratory. Older participants received monetary compensation for their time and travel expenses, and younger participants received course credit.

Measures

All participants completed a series of tests to assess traditional neuropsychological abilities. The following measures were administered individually in a single 2-hr session, and scored by trained research assistants according to standardized procedures. The testing order for the entire session is available upon request.

Sandbox Task

We modified the between-subject and within-subject procedures outlined by Sommerville et al. (2010). Specifically, participants were tested using a rectangular sandbox constructed from pine (152 cm long by 45 cm wide by 31 cm deep). The sandbox was filled with Styrofoam peanuts to 2.54 cm below the top lip. The sandbox was painted grey with a green base and sat upon a table. Participants stood in front center of the sandbox. The experimenter stood directly across from participants on the other side of the sandbox.

The experimenter, unaware of the experimental hypotheses, narrated a story, and placed objects within the sandbox while participants watched. There were three within-subject conditions representing trial type: In the false belief condition and the memory control condition, a protagonist put an object in one location, and while the protagonist was absent, a second character moved that object to a new location. After a short delay, participants were asked to indicate where the protagonist would look for the object upon her return (false belief) or where the protagonist originally placed the object before leaving (memory control). In the no false belief condition, a protagonist put an object in one location, and while the protagonist was absent, a second character subsequently placed another object in a new location.

For example, in one story participants were told, "Sally and Ann are outside playing in the sandbox. Sally hides a red toy dog in the sand here (experimenter hides red toy dog at the first hiding location) and then goes inside to get a drink of water." The procedure then varied according to condition. Participants in the false belief and memory control conditions next were told: "While Sally is inside the house, Ann finds the toy dog and hides it here," and the experimenter hid the toy dog at the second hiding location. Participants in the no false belief condition next were told, "While Sally is inside the house, Ann takes a lego piece and hides it here," and the experimenter hid the lego piece at the second hiding location. Participants in the false belief and no false belief condition next were asked the critical question, "When Sally comes back outside to play, where is she going to look for her red toy dog?" Participants in the memory control condition, identical to the false belief condition up to the critical question, next were asked: "Then Sally comes back outside to play. Where did she hide her red toy dog before she left?" The no false belief and memory control conditions control for placement of an object at a second location in the sandbox and memory for the original placement of the object, respectively. Both control conditions place task demands on the participant similar to those in the false belief condition, except that the control conditions can be solved without ignoring one's own privileged knowledge to reason about an actor's false belief.

Each trial involved a different story. Thus, the box represented a different continuous hiding location on each trial (garden plot, sandbox, deep chest freezer, bath tub, forest, and street) and featured a different set of characters and different hiding objects (hiding objects were miniature replicas of the objects mentioned in the story; see http://sandboxtask.weebly.com/for the stories).

There were six trials in total consisting of two trials for each of the three different trial types (false belief, no false belief, memory control). Three different versions of the six sandbox scenarios were created. The six scenarios and object placements were presented in a fixed order. The three trial types were presented in counterbalanced order within each age group. On all trials, 36 cm separated the first and second hiding locations. One trial of each type involved a second hiding location to the left of the first hiding location, whereas the other trial involved a second hiding location to the right of the first hiding location. Each trial contained a different set of hiding locations. In between the story and the critical question, participants completed a 45-s visual distractor task (Where's Waldo), which we administered to prevent them from using perceptual strategies to guide their search (e.g., by continuing to fixate on the first hiding location). The experimenter also leveled the Styrofoam peanuts during this time to ensure that participants could not use surface cues to guide their search estimates. After 45 s participants indicated where the protagonist would look for the object when he or she returned by pointing to the exact location on the surface of the Styrofoam peanuts. At the end of the testing procedure, the experimenter used a ruler to score participant's search responses on the task, using a set of concealed stickers on the experimenter's side of the box that indicated the participants' responses on each trial. A bias score was calculated for each trial by measuring the distance away from the first location toward the second location. This could result in negative, zero, or positive bias scores. The two bias scores for each trial type were averaged to yield an average bias score for the three trial types. Lower scores approaching zero indicate less bias, and by extension, better performance. Higher scores indicate more bias.

Verbal Memory

We measured verbal memory with the California Verbal Learning Test-II (CVLT-II). The CVLT-II (Delis, Kramer, Kaplan, & Ober, 2000) assesses learning over repeated trials, susceptibility to memory interference, and delayed verbal memory. Participants were presented with 16 standardized items over five trials. Recall was assessed after each trial, after an interference trial, and again after a 20-min delay. The Long Delay Free Recall provides an estimate of the amount of verbal information an individual is able to retain after a delay of approximately 20 min, and performance on this measure was used to index memory.

Executive Function

The Trail-making, Color-Word Interference, and Verbal Fluency subtests of the Delis-Kaplan Executive Functioning System (Delis, Kaplan, & Kramer, 2001) were used to examine three aspects of executive functioning: mental set shifting, cognitive inhibition, and response monitoring, respectively (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). For the Trail-making test, we used the commonly reported

Trails Test B as the dependent variable. The amount of time required to connect numbers and letters correctly in an alternating pattern is the score for this measure. The Color-Word Interference subtest (also known as the "Stroop" task) requires inhibition of a dominant response (reading) in favor of a less dominant response (color-naming). An interference score is derived by subtracting the baseline (color-naming) condition from the higher-level (inhibition) condition. The Verbal Fluency subtest requires the generation of as many items beginning with certain letters (e.g., "S") as possible within a specified time frame.

Processing Speed
We assessed processing speed with the Digit Symbol–coding subtest of the Wechsler Adult Intelligence Scale—III. For this task, the participant is presented with a series of numbers paired with symbols, and is required to sequentially draw the paired symbols within the designated squares as quickly as possible. The total number of correctly copied symbols served as the raw score for processing speed.

Working Memory
We assessed working memory with the Letter-Number Sequencing subtest from the Wechsler Adult Intelligence Scale—III. For this task, the examiner read different combinations of numbers and letters and the participant was asked to first recall the number(s) in ascending order followed by the letter(s) in alphabetical order. In order to advance to the next item, the participant had to complete at least one of the three trials within the current item successfully. The total number of correctly recalled items served as the raw score for working memory abilities. We added this task after we had started testing, and therefore, only a subset ($n = 57$) of our sample completed this task.

Vocabulary
To index general knowledge of word meanings, we administered both Vocabulary (V-2) and Extended Range Vocabulary (V-3) from the ETS kit (Ekstrom, French, Harman, & Derman, 1976). These multiple-choice tests require participants to select a word closest in meaning to the target word from a list of five choices. The total score reflects the number marked correctly, minus a fraction (.25) for each marked incorrectly to discourage guessing.

Statistical Analysis

Group differences across the demographic and cognitive variables were examined with one-way analyses of variance (ANOVAs). Distributions were examined within each age group for extreme values and departures from normality on the Sandbox task to insure appropriateness in parametric testing. Fifteen data points (out of 282 possible, or 5% of the data) emerged as extreme outliers (>3 *SD*) for the following trial types: false belief (younger = 6; middle-aged = 2; older = 0), no false belief (younger = 1; middle-aged = 1; older = 1), memory control (younger = 2; older = 2). These values were adjusted, rather than omitted, to make them more contiguous with the data while maintaining their distal-most ranking in the distribution (Tabachnick & Fidell, 2007). Thus, extreme scores were "brought in" to the normal distribution, and assigned scores slightly higher than the highest scores or slightly lower than the lowest scores within the normal distribution. Note that we removed one younger participant from further analysis, because his anomalous performance led to concerns that he did not understand the task demands.

Because detailing neuropsychological functioning was not a focus of the current study, we conducted a principal component analysis with the following seven neuropsychological test scores to reduce the number of dependent variables: CVLT-II Long Delay Free Recall total, CVLT Immediate Recall, Trails B completion time, Stroop interference score, Processing Speed, Verbal Fluency, and ETS Vocabulary total score. This analysis revealed the presence of two components with eigenvalues exceeding 1.0, explaining 44% and 18% of the variance, respectively. The first component included measures of processing speed, learning and memory, mental set shifting, and cognitive inhibition. We henceforth refer to this as the Neuropsychological Ability component. The second component consisted of Vocabulary and Letter Fluency measures. We refer to this as the Word Knowledge component. We retained subjects' factor scores on these two components for use in subsequent analyses.

To address our first prediction, we conducted a repeated measures analysis of variance utilizing a 3 (belief) by 3 (age group) mixed linear model, in which belief (false belief, no false belief, memory control) was a within-subject factor and age group (younger, middle-aged, older) was a between-subject factor. To address our second prediction, we conducted correlations among our main variables of interest for all age groups combined and for each age group separately. These can be seen in Table 1. We then examined the overall (age group combined) correlations between the cognitive variables and false belief bias. For each variable that correlated significantly with false belief bias, we added this as a predictor variable, along with age group in a multiple regression with false belief bias as the dependent variable. Analyses were conducted using SPSS 16 software (SPSS, Chicago, IL).

Table 1. Zero-order correlations

	Neuropsych.	Language	LNS	False belief	No false belief	Memory control
		Age groups combined				
Education	−.25*	.31*	−.15	.15	.02	.01
Neuropsych.		.00	.27*	−.23*	−.31*	−.15
Language			.22	.14	−.19	−.14
LNS				−.11	−.14	.04
False belief					.12	−.16
No false belief						−.03
		Younger adults				
Education	.00	−.05	−.23	.27	−.08	−.07
Neuropsych.		.20	−.17	.11	.03	−.11
Language			.11	−.33*	−.26	−.04
LNS				−.19	.17	.09
False belief					−.17	−.13
No false belief						.02
		Midle-aged adults				
Education	.18	.14	.03	.35	.07	−.25
Neuropsych.		.35	.64**	−.10	−.37	.07
Language			.35	.12	−.20	−.34
LNS				.31	−.61	.48
False belief					.33	−.06
No false belief						−.22
		Older adults				
Education	−.25	.29**	−.05	−.09	−.00	.10
Neuropsych.		.44*	.80*	.03	−.53*	−.11
Language			.58	.14	−.27	−.18
LNS				.37	−.60**	−.30
False belief					.06	−.34*
No false belief						−.08

Note. Neuropsych. = Neuropsychological ability component score; Language = Word knowledge component score; LNS = Letter-Number Sequencing (completed by a subset of participants, $N = 57$).

*$p < .05$; **$p < .10$.

RESULTS

Participant Characteristics

Information regarding demographic characteristics and mean performance on the variables of interest is presented in Table 2. Not surprisingly, given the truncated educational attainment of the young adult sample (all were still enrolled in university), the middle-aged participants were more educated than either the younger or older

Table 2. Demographic and task means (standard errors)

Variable	Younger adults (a)	Middle-aged adults (b)	Older adults (c)	Significant differences
Age	19.22 (.23)	56.30 (.52)	67.60 (.99)	a < b < c
Female/Male	28/9	16/4	28/9	ns
Education	13.22 (.20)	15.10 (.46)	14.35 (.46)	a < b, c
CVLT long delay free recall*	13.25 (.34)	10.95 (.60)	10.38 (.58)	a > b, c
CVLT—Immediate Recall*	7.53 (.32)	6.35 (.44)	5.38 (.25)	a > b, c
Trails B time#	52.32 (2.07)	65.00 (3.62)	84.32 (5.94)	a, b < c
Stroop time#	43.24 (1.31)	50.90 (2.06)	56.54 (1.69)	a < b < c
Digit Symbol*	93.22 (2.12)	78.05 (2.36)	67.00 (2.22)	a > b > c
Vocabulary*	14.68 (.84)	30.22 (1.46)	26.86 (1.36)	a < b, c
Verbal Fluency*	41.16 (1.42)	45.20 (2.79)	43.14 (1.76)	*ns*
Letter-Number Sequencing*	11.92 (.34)	12.62 (.88)	10.11 (.75)	a, b > c
False belief bias#	3.08 (1.80)	8.84 (2.45)	11.70 (1.80)	a < b, c
No false belief bias#	1 .47 (.78)	1.74 (1.06)	3.18 (.78)	*ns*
Memory control bias#	2.94 (1.11)	1.69 (1.51)	5.22 (1.11)	*ns*

Note. *Higher scores reflect better performance.
#lower scores reflect better performance. Significant differences = follow-up planned comparisons ($p < .05$) conducted after a significant one-way ANOVA; a < b, c = younger group significantly less than middle-aged and older group, which do not differ from each other.

adults. As expected, younger adults generally outperformed older adults on the measures of processing speed, learning/memory, and executive functioning, and exhibited less word knowledge than middle-aged or older adults. The gender composition was similar in all age groups.

False Belief

Consistent with our predictions, a main effect emerged for belief $F(2, 182) = 11.84$, $p < .001$, $\eta_p^2 = .11$, indicating that bias was greatest in the false belief condition ($M = 7.87$, $SE = 1.18$) compared to the no false belief ($M = 2.13$, $SE = .51$) and memory control ($M = 3.29$, $SE = .72$) conditions. Furthermore, a main effect emerged for age group $F(2, 91) = 9.33$, $p < .005$, $\eta_p^2 = .17$, indicating that older adults exhibited more biased performance overall ($M = 6.70$, $SE = .69$) than the middle-aged adults ($M = 4.09$, $SE = .94$), who showed more bias overall than younger adults ($M = 2.50$, $SE = .69$).

Importantly, these effects were qualified by a significant Belief × Age Group interaction, $F(4, 182) = 2.45$, $p = .05$, $\eta_p^2 = .05$ (see

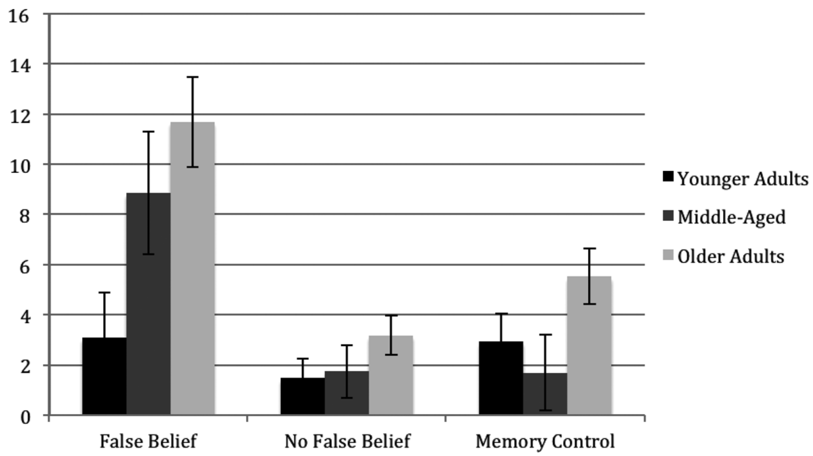

Figure 1. Mean bias score in cm for false belief, no false belief, and memory control conditions in the Sandbox task. A score of 0 indicates no bias toward the second location; a positive score indicates bias toward the second location. Error bars are standard errors.

Figure 1).[1] To explore this interaction further, we conducted three separate one-way ANOVAs on each of the three levels of belief, using age group as the between-subject factor. This resulted in a significant effect of age group for false belief $F(2, 93) = 5.85$, $p < .01$, but not for no false belief $F(2, 93) = 1.33$, $p = .27$, or memory control $F(2, 93) = 2.04$, $p = .14$. Follow-up planned contrasts revealed that younger adults ($M = 3.08$, $SEM = 1.23$) showed less false belief bias than middle-aged adults ($M = 8.84$, $SEM = 2.83$), $t(55) = 2.16$, $p = .03$, and older adults ($M = 11.70$, $SEM = 2.10$), $t(55) = 3.54$, $p = .001$. Middle-aged and older adults did not differ in terms of their false belief bias ($t < 1.0$).

We then addressed our second research question, which was whether age differences in false belief bias were modified by performance in other cognitive domains. Toward this end, we first examined the correlations between potential cognitive variables and belief. As can be seen in the top section of Table 1, neuropsychological ability correlated significantly with false belief bias and no false belief bias when we considered all age groups combined.

[1]Mauchly's sphericity was significant, indicating that the variances associated with our different age groups on the dependent variable differed. In fact, in the two older adult age groups, the variances in the false belief condition were nearly twice those of the no false belief and memory control conditions. Correcting for the violation of sphericity using Huynh-Feldt changed the interaction only slightly $F(3, 141) = 2.45$, $p = .06$.

However, as can be seen in the three bottom sections of Table 1, these correlations largely disappeared when we considered each age group separately. To test more conservatively whether neuropsychological ability accounted for the age group differences in false belief bias that we observed, we conducted a multiple regression with false belief bias as the dependent variable and age group and neuropsychological ability entered in succession as predictor variables. Only age group accounted for significant variance in false belief bias (R^2_{change} = .11, p = .001). Neuropsychological ability failed to account for any additional variance in false belief bias (R^2_{change} = .00, p = .89). These results show that neuropsychological ability did not account for the age-related differences in false belief bias that we observed.[2]

DISCUSSION

We began this study with a simple question: How does false belief understanding, an important facet of mental state reasoning and ToM, change from younger to older adulthood? We administered a continuous false belief task that was designed to measure mental state reasoning across the life span (Sommerville et al., 2010); however, we focused here on younger, middle-aged, and older adults. As expected, middle-aged and older adults exhibited more false belief bias than did younger adults. Importantly, this difference was not accounted for by several cognitive variables that have been shown in previous work to account for much of the difference between younger and older adults on ToM tasks, namely language ability, executive function, processing speed, and memory (Bailey & Henry, 2008; Bull et al., 2008; German & Hehman, 2006; Sullivan & Ruffman, 2004).

What is unique about the current study is that we used a continuous measure of false belief (the Sandbox task) to distinguish mental state reasoning performance among younger, middle-aged, and older adults. Most previous work has focused on younger and older adults

[2]Because Location A was the correct location in all three trial types in the Sandbox task, participants could have learned this fact over subsequent trials and thereby shown reduced bias scores over trials. We tested for this possibility by conducting a repeated measures ANOVA with trial number (1–6) as the within-subject factor and age group (younger, middle-aged, older) as the between-subject factor. This resulted in a main effect for trial number F(5, 86) = 3.27, p = .01, $\eta_p^2 = .16$,, and age group F(2, 90) = 3.86, p = .02, $\eta_p^2 = .08$. The interaction was not significant (p = .39). Further exploration of the trial number main effect, the effect of interest here, revealed that bias scores fluctuated over trials for all three age groups, but did not decline over trials in any of the age groups. This result clearly shows that, over trials, participants failed to learn that Location A was the correct location.

only (60 years of age and older). One previous study included younger, middle-aged, and older adults and showed that emotion understanding, an important facet of mental state reasoning, begins to decline in middle age (Pardini & Nichelli, 2009). We replicated and extended this data pattern here by generalizing the age-related decline in ToM to a new perspective-taking false belief task.

Although theories abound as to why developmental changes to ToM abilities occur in childhood (see Wellman et al., 2001, and commentaries), we are left with few theories to explain why an apparent decline in these abilities emerges in midlife and continues through older age. Age-related declines are well-documented on more "fluid" aspects of cognition such as processing speed, working memory, and executive function (Baltes, 1987; Craik & Bialystok, 2006). These abilities have been linked to performance on false belief and other ToM tasks in children as well as younger and older adults. To the extent that mental state reasoning abilities are compiled from these basic mental abilities that tend to decline with age, related declines in mental state reasoning should also emerge. Consistent with this view are previous findings demonstrating that age-related declines on certain ToM tasks are mediated in part by declines in "fluid" aspects of domain-general cognitive functioning (Maylor et al., 2002; Sullivan & Ruffman, 2004).

Nonetheless, some prior work also described age-related changes to ToM that are independent of domain-general cognitive abilities such as executive function, processing speed, and memory (Sullivan & Ruffman, 2004). Similarly, our findings suggest that although performance on such domain-general measures is associated with false belief bias, the age differences observed are independent of these cognitive abilities.

Notably, we did not manipulate cognitive load and executive function *while* participants completed the Sandbox task in the present study. Other work has demonstrated the utility of directly manipulating cognitive load in younger (Bull et al., 2008) and older participants while they complete ToM tasks (McKinnon & Moscovitch, 2007). Future work should explore the effects of dual tasks and cognitive load on performance using a battery of tasks believed to tap different aspects of ToM, including the Sandbox task. This approach could be used with adult and aging populations, as well as children.

A cautionary note about using a battery of tasks to tap different aspects of ToM is in order. Tasks believed to measure the same underlying construct (e.g., inhibition) do not always correlate with each other (see Friedman & Miyake, 2004). Similarly, tasks believed to measure ToM in adults (e.g., The Mind in the Eyes task, the Stories task) do not always correlate with each other (Brent, Rios, Happe, & Charman, 2004; Roeyers, Buysee, Ponnet, & Pichal, 2001). A general problem facing researchers wishing to develop new measures of an established construct is that they must demonstrate that their new measure correlates with existing measures. If the existing measures correlate modestly with each other, then the new measure will likely only correlate modestly with these other measures. Inconsistent and weak correlations reported in the literature among ToM tasks, especially adult ToM tasks, indicate that these tasks are not only measuring different aspects of ToM, but that ToM is not a unitary construct.

Assuming that ToM is a collection of different cognitive and social abilities, we maintain that inconsistencies in the literature with respect to what underlies adult age differences in ToM could be due to the specific component processes required to perform different ToM tasks. For example, Bailey and Henry (2008) used a task developed by Samson and colleagues (the reality-unknown task) to measure false belief deficits that require different levels of self-perspective inhibition. The reality-unknown task manipulates inhibitory demands across different experimental conditions (low versus high inhibition). Bailey and Henry found that older adults outperformed younger adults on the high-inhibition version of the reality-unknown task; however, performance on the Stroop task, a measure of inhibition, accounted for this age-related difference in ToM. The Sandbox task, in contrast to the reality-unknown task, contains various levels of inhibitory demand, but it does not manipulate these demands directly. Moreover, in the present study, inhibition, also measured by the Stroop task, correlated with false belief performance on the Sandbox task; however, inhibition did not account for the age difference observed on the false belief condition of the Sandbox task. One reason for these results is that the Sandbox and reality-unknown tasks likely tap different aspects of ToM. Future work should explore the specific component processes underlying different adult ToM tasks.

German and Hehman (2006) argue for using what is called the minimal pair design in which the experimenter manipulates content between two conditions in an experimental task via the smallest change possible. Such an experimental design permits more precise interpretation of task-relevant performance differences. The Sandbox task can be construed as using a minimal pair design in that the three belief conditions reported in the current study differ either by slight wording or slight changes to the story.

Given the structure of the Sandbox task, one may wonder what exactly it measures. The Sandbox task taps false belief (and by extension, ToM) errors on a continuum. The task measures subjects' bias scores toward the second location in three different belief conditions (false belief, no false belief, memory control). By "bias," we mean the relative "pull" away from the correct location toward the incorrect location in the box. If this pull differs across the three belief conditions and across the adult life span, as it does, we conclude that younger adults show less false belief bias than do middle-aged and older adults.

Our findings also suggest that models for how an understanding of beliefs develops might be reconsidered. Most of the developmental literature on ToM depicts children's understanding of beliefs as developing in a discontinuous, categorical fashion: Children either have the concept of belief, or they do not. From this perspective, once this understanding is acquired, adults can consistently use a person's belief state to predict that person's future actions, even when that belief state conflicts with reality. Thus, the development of understanding beliefs is much like a light switch that can go from being set in the off position (e.g., no understanding that beliefs can conflict with reality), to being set in the on position (e.g., understanding that beliefs can conflict with reality).

Given our findings, as well as those of other laboratories (Apperly, Samson, & Humphreys, 2009), we suggest that false belief reasoning may follow a more continuous, graded, U-shaped development. We argue that false belief tasks have two primary components: the ability to represent another's belief state, and the ability to ignore the pull of reality and/or one's own privileged knowledge. Developmental change occurs as a function of the relative dominance and contribution of these two abilities. From this perspective, false belief reasoning is more like a dimmer switch, with one end of the switch representing reality, and the other end of the switch representing a particular belief state. Our Sandbox task permits fine-grain assessment of the *relative* amount by which people are biased by their own privileged knowledge of reality when estimating what a naïve person knows.

The current findings add to a growing body of research indicating that social reasoning, everyday problem solving, cognitive empathy, and social functioning show reliable declines with increasing age (Pratt, Diessner, Pratt, Hunsberger, & Pancer, 1996; Thornton & Dumke, 2005). Although it may be presumed that empathy and mental state reasoning are required for successful social functioning, it will be important for future research to investigate the functional implications of these findings. Furthermore, the cross-sectional nature of this and other studies limits our ability to draw causal inferences. Future longitudinal investigations may shed additional light on important questions regarding the mechanisms underlying the age differences in ToM that we observed.

CONCLUSION

False belief errors persist throughout life, and are magnified in preschoolers and again in later adulthood. In some studies, these differences in false belief understanding and mental state reasoning appear to be due in part to differences in executive function (Bull et al., 2008; German & Hehman, 2006; McKinnon & Moscovitch, 2007). In the current study, middle-aged and older adults showed more false belief difficulty than did younger adults; however, these age-related differences in ToM were not due to age group differences in language ability, executive functioning, processing speed, or memory. We therefore conclude that ToM, as measured by a continuous false belief task, decreases from younger to older adulthood irrespective of age-related changes to other aspects of cognitive functioning.

REFERENCES

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841–844.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45,* 190–201.

Bailey, P. E., & Henry, J. D. (2008). Growing less empathic with age: Disinhibition of the self-perspective. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 63,* P219–P226.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamincs between growth and decline. *Developmental Psychology, 23,* 611–626.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21,* 37–46.

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 and Allied Disciplines, 42,* 241–251.

Bernstein, D. M., Atance, C., Loftus, G. R., & Meltzoff, A. (2004). We saw it all along: Visual hindsight bias in children and adults. *Psychological Science, 15,* 264–267.

Bernstein, D. M., Atance, C., Meltzoff, A. N., & Loftus, G. R. (2007). Hindsight bias and developing theories of mind. *Child Development, 78,* 1374–1394.

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18,* 382–386.

Brent, E., Rios, P., Happe, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8,* 283–299.

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition, 107,* 663–672.

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

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

Charlton, R. A., Barrick, T. R., Markus, H. S., & Morris, R. G. (2009). Theory of mind associations with other cognitive functions and brain imaging in normal aging. *Psychology and Aging, 24,* 338–348.

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences, 10,* 131–138.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS).* San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California verbal learning test* (2nd ed.). San Antonio, TX: Psychological Corporation.

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests* (Rev. Ed.). Princeton, NJ: Educational Testing Service.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control conditions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133,* 101–135.

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.

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

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.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25–41.

Kukull, W. A., Larson, E. B., Teri, L., Bowen, J., McCormick, W., & Pfanschmidt, M. L. (1994). The Mini-Mental State Examination score and the clinical diagnosis of dementia. *Journal of Clinical Epidemiology, 47,* 1061–1067.

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 performance on theory of mind tasks decline in old age?. *British Journal of Psychology, 93*(Pt 4), 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.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78,* 622–646.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640.

Pardini, M., & Nichelli, P. F. (2009). Age-related decline in mentalizing skills across adult life span. *Experimental Aging Research, 35,* 98–106.

Pratt, M. W., Diessner, R., Pratt, A., Hunsberger, B., & Pancer, S. M. (1996). Moral and social reasoning and perspective taking in later life: A longitudinal study. *Psychology and Aging, 11,* 66–73.

Roeyers, H., Buysee, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of Child Psychiatry and Psychology, 42,* 271–278.

Slessor, G., Phillips, L. H., & Bull, R. (2007). Exploring the specificity of age-related differences in theory of mind tasks. *Psychology and Aging, 22,* 639–643.

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2010). Measuring false belief in centimeters. Manuscript under revision.

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

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.

Thornton, W. J., & Dumke, H. A. (2005). Age differences in everyday problem-solving and decision-making effectiveness: A meta-analytic review. *Psychology and Aging, 20,* 85–99.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

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

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.