

## Research Article

**Cite this article:** Yow, W.Q. and Li, X. (2025). Theory-of-mind understanding in aging: Effects of early bilingual language experience. *Bilingualism: Language and Cognition* 1–13. <https://doi.org/10.1017/S1366728925000240>

Received: 03 December 2023

Revised: 01 January 2025

Accepted: 10 February 2025

### Keywords:


aging; theory-of-mind; bilingualism; second language age-of-acquisition

### Corresponding author:

W. Quin Yow;

Email: [quin@sutd.edu.sg](mailto:quin@sutd.edu.sg)

Both authors contribute to this manuscript equally.

 This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

## Abstract

The ability to understand and speak more than one language (i.e., bilingualism) may provide benefits to preserving social cognition against normal age-related deteriorations. This study examined how variations in bilingual language experience influence theory-of-mind (ToM) understanding in late adulthood. One hundred and five cognitively healthy older adults ( $M_{\text{age}} = 66.23$  years, range = 56–79) and 80 young adults ( $M_{\text{age}} = 22.03$  years, range = 19–30), who were bilingual speakers, completed a ToM task battery, a self-report questionnaire on their language background, and a battery of general cognition assessments. We found an overall age-related decline in ToM, where older adults made more errors in inferring others' mental states compared with young adults. Importantly, an earlier L2 age of acquisition (L2AoA) predicted better ToM performance among older adults, over and above the effects of age, education, and general cognition. The results suggest that early bilingual acquisition may enhance social cognitive processes during development and contribute to intact ToM abilities in older adult bilinguals.

## Highlights

1. Older adults made more errors in theory-of-mind (ToM) tasks than younger adults.
2. Earlier L2AoA predicted enhanced ToM performance in older bilinguals.
3. This effect was over and beyond the effects of age, education, and cognitive ability.
4. Bilingual usage, proficiency, and code-switching showed no significant effects on ToM.
5. Results indicate early bilingualism may protect against age-related ToM decline.

## 1. Introduction

To navigate the social world and participate in positive social interactions, theory-of-mind (ToM)—the ability to understand the mental states of others independently from our own—is necessary for understanding the behaviors and needs of others across the lifespan (Baron-Cohen et al., 1993; Caputi et al., 2012; Frith, 2008). Despite some mixed evidence, many recent studies have shown that older adults are susceptible to a decline in ToM in normal aging (for reviews and meta-analyses, see Fernandes et al., 2021; Henry et al., 2013; Roheger et al., 2022). Importantly, while evidence shows that bilingual experience may be beneficial to ToM development in childhood and young adulthood (e.g., Navarro & Conway, 2021; Nguyen & Astington, 2014; Rubio-Fernandez & Glucksberg, 2012; Schroeder, 2018), research examining bilingual effects on ToM in the aging population is limited (see Feng et al., 2023). In this article, we seek to examine the relationship between bilingual experience and ToM performance in late adulthood and explore the role of bilingualism in contributing to intact ToM ability against normal age-related deteriorations.

Traditionally, the ToM literature has focused on children, where there is largely a consensus that ToM ability develops in early childhood (Wellman & Liu, 2004) and continues to improve throughout adolescence (Dumontheil et al., 2010; Osterhaus & Koerber, 2021), with performance peaking in young adulthood (Osterhaus & Bosacki, 2022). For instance, by 3 years old, children demonstrate some early ability to detect and identify others' emotions and desires (e.g., Pons et al., 2004). By age 5, children develop their basic ToM understanding and can explicitly reason about others' mental states that are different from their own, such as perspective-taking and first-order false belief (e.g., Wellman et al., 2001). From age 6 to the teen years, children become more skilled at inferring complex emotional and mental states, such as higher-order emotions and false beliefs, lies, sarcasm, metaphor, and so on (e.g., Osterhaus et al., 2016; Warnell & Redcay, 2019).

In recent years, there has been an increasing number of research that attempted to understand how ToM ability changes beyond young adulthood (e.g., Bradford et al., 2023; Grainger et al., 2023; Krendl et al., 2023; Wang & Su, 2013). With a few exceptions (e.g., Dodell-Feder et al., 2020; Happé et al., 1998), current findings suggest an age-related decline in ToM ability, typically starting around 40–50 years of age (e.g., Bradford et al., 2023; Grainger et al., 2023; see Fernandes et al., 2021; Henry et al., 2013, for reviews). For example, using the Director Task where

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

participants must consider the perspective of the “director” to infer the referential target, Bradford et al. (2023) reported that middle-aged (41–62 years old) and older (63–86 years old) adults made increasing errors with advancing age across middle and later adulthood, whereas young adults (20–40 years old) performed at ceiling levels in this task. There has been evidence showing age-related differences in ToM as a function of cognitive demands, where cognitive ability is often found to decline with healthy aging (e.g., Phillips et al., 2011; Rahman et al., 2021). However, a recent meta-analysis by Roheger et al. (2022) revealed that healthy older adults (defined as 50 years old and above) performed significantly worse than young adults (below 50 years old) in all ToM aspects identified in their analysis, including cognitive ToM (i.e., inferences about thoughts, intentions, and beliefs), affective ToM (i.e., inferences about feelings), and mixed ToM (i.e., cognitive and affective ToM not differentiated), with the largest effect size for cognitive ToM.

Thus, while the findings seem consistent in showing a general decline in older adults’ ToM, meta-analyses have indicated a certain level of variability in the reported effects of age in past studies (e.g., Fernandes et al., 2021; Roheger et al., 2022). This variability may be attributable to methodological limitations in the literature, such as the use of different tasks purported to measure different aspects of ToM and different populations with varying age cutoffs and education backgrounds (e.g., Henry et al., 2013; Raimo et al., 2022). For example, Roheger et al. (2022) found that studies using cognitive ToM tasks, such as first- and second-order false-belief tasks and the Director Task, generally reported a larger effect size (i.e., greater difference between younger and older adults) than studies using affective ToM tasks, such as facial recognition tasks and the Reading the Mind in the Eyes Tests (RMET). Older adults also appeared to show declined performance only on more complex ToM tasks that required higher-order mental inferences (e.g., McKinnon & Moscovitch, 2007). These findings highlight the importance of considering task specificity when examining age-related effects in ToM research.

Moreover, while many studies compared the ToM performance of people from different age groups (e.g., most often young adults versus older adults, sometimes including a third group of middle age), the age cutoffs for these groups are often arbitrary and vary across studies. A meta-analysis by Rahman (2021) highlighted this variability, showing that the mean reported age of the older adult group ranged from 65–84 across 28 studies. The analysis found a medium age effect for studies with a younger group of older adults (range of mean age: 60–74) and a large age effect for studies with an older group of older adults (mean age 75 and above). This variability underscores a key challenge in aging research: using age groups with arbitrary cutoffs may affect the interpretation of findings and limit our understanding of the trajectory of age-related decline throughout adulthood. Despite these limitations, the categorical approach remains widely used as it allows for direct comparisons between older adults (within a defined age range) and younger adults, who are often assumed to be at their peak cognitive performance. In this study, we compared older and younger adults using age groups while also treating age as a continuous variable within each group of adults to examine how bilingual language experience influences their ToM performance, which tends to decline after midlife. However, we acknowledge recent efforts to explore alternative approaches, such as sampling participants across a broader age range from young to old (e.g., Bradford et al., 2023). We will discuss this issue in greater detail in our analysis and interpretation of the data.

Fundamentally, some researchers have argued that, when studying adult ToM, its operational definition (including its cognitive and affective component) and assessment are often “vague and inconsistent” (Schaafsma et al., 2015, p. 65; also see Quesque & Rossetti, 2020; Yeung et al., 2024), and they have called for research with more specific hypotheses regarding ToM in the topics of investigation. Given that our goal is to examine the effects of bilingualism on ToM abilities in young and older adults, we employ the classic definition of ToM as the ability to attribute mental states (such as beliefs, intentions, and emotions) to others (Frith & Frith, 2006; Gallese & Sinigaglia, 2011). Thus, our ToM assessments include belief, intention, and emotion inferences across various levels of complexity (e.g., from basic to advanced ToM judgments). We aim to provide evidence on the impact of age and bilingual language experience on general ToM abilities in the aging process.

We propose that there is significant heterogeneity in age-related changes in ToM across individuals, even within the common aspects of ToM alterations (e.g., Grainger et al., 2023; Greenberg et al., 2023; Otsuka et al., 2024; Roheger et al., 2022). The rate and extent of these age-related changes in ToM may be influenced by individual experiences that could account for longitudinal changes in social cognition and the brain (e.g., Dotson & Duarte, 2020; Henry et al., 2023; Li et al., 2024). We draw parallels with the revised Scaffolding Theory of Aging and Cognition (STAC-r; Reuter-Lorenz & Park, 2014), which posits that life-course factors (e.g., experience and environment) influence the developmental course of brain structure, function, and cognition over time. These life-course factors, such as education and social engagement, may increase cognitive reserve, the brain’s ability to actively use other brain pathways to compensate for age- or disease-related deterioration (Stern, 2002; 2009). Consequently, these experiences allow individuals to better cope with cognitive challenges and mitigate the adverse effects of age-related structural and functional brain changes (Stern et al., 2019). For example, middle-aged and older adults who engaged more actively in cognitive, leisure, and social activities were reported to have better cognitive ability, less age-related cognitive decline, and a reduced risk of dementia diagnosis than their less-active peers (e.g., Wilson et al., 2013).

Importantly, variables such as bilingualism have been identified as having a positive impact on cognitive aging (e.g., Bialystok, 2021; Gallo & Abutalebi, 2023). Both theoretical and empirical work suggest bilingualism contributes to cognitive and brain reserve, helping to protect against cognitive decline with age, with either preserved brain functional efficiency or greater neural tissue in bilingual older adults compared with their monolingual counterparts (e.g., Abutalebi et al., 2014; Gallo et al., 2021; Luk et al., 2011; Stevens et al., 2023; see Anderson et al., 2020; Gallo et al., 2022, for reviews). Remarkably, in cases where bilinguals show accelerated brain atrophy, their cognitive performance remains relatively preserved (Anderson et al., 2021; Perani et al., 2017; Schweizer et al., 2012). Individual differences in brain functional correlates are also associated with a wide range of bilingual experiences, including onset age of bilingualism (e.g., DeLuca et al., 2020; Luk et al., 2020), bilingual language proficiency and usage (e.g., DeLuca & Voits, 2022; Sulpizio et al., 2020), and language diversity (i.e., language entropy, e.g., Gullifer et al., 2018; Li et al., 2021). Thus, it has been suggested that bilingualism offers protection against cognitive decline via compensatory scaffolding, such that it can provide additional support to compensate for declining neurocognitive function with age.

Based on the STAC-r framework, we postulate that life-course enrichment factors, such as individual language experience in a

bilingual context, would similarly offer enhancement and compensatory scaffolding to *social cognitive ability* during aging. Previous research on children and younger adults suggests a link between bilingualism and ToM advantages. A large body of research has reported that bilingual children outperformed monolingual children in ToM-based tasks (e.g., Fan et al., 2015; Kovács, 2009; see Schroeder, 2018, for a meta-analysis). There is also evidence that bilinguals' superior ToM performance in childhood continues into young adulthood (e.g., Lorge & Katsos, 2019; Navarro & Conway, 2021; Rubio-Fernandez & Glucksberg, 2012; cf. Ryskin et al., 2014). For example, Navarro & Conway (2021) revealed better perspective-taking ability in bilingual versus monolingual college students (mean age = 27.29 years). Going beyond comparisons between monolinguals and bilinguals, two studies examined whether individual differences in bilingual experience influence perspective-taking or mentalizing skills among young adult bilinguals (e.g., aged 18–35; Navarro et al., 2022; Tiv et al., 2021). It was found that more regular L2 usage and greater childhood exposure to diverse languages both contributed to better perspective-taking ToM in a sample of bilingual participants (mean age = 25.27 years), but L2 fluency or number of languages spoken did not have an impact on their ToM performance (Navarro et al., 2022). The effects of bilingualism on ToM may arise from certain aspects of the bilingual experience that foster early development of sociolinguistic sensitivity (i.e., the awareness of other people's mental states). For example, encountering people from diverse language backgrounds may "train" bilinguals to infer what other people know and do not know (also see Yu et al., 2021).

The effects of bilingual experience on older adults' ToM, however, have not yet been well studied. To our knowledge, only one study has administered a ToM task (i.e., the Faux Pas task, "a measure of complex ToM"; p. 299) together with six cognitive tests in monolingual and bilingual participants at age 74 as part of a longitudinal study (Cox et al., 2016). They found some weak evidence for a bilingual advantage on the Faux Pas task, but this advantage was attenuated when controlling for pre-existing differences in childhood intelligence and social class. Although informative, this study mainly focused on more advanced ToM and did not sufficiently capture individual variability in bilingual experience (e.g., participants who could communicate in a second language were coded as bilingual). Our study seeks to examine the relation between aging, bilingual experience, and ToM (with a variety of content and complexity), which could provide important insights into how such life-course enriching experience (e.g., early acquisition of a second language or using more than one language) on ToM skills in later life. Lifelong bilingualism may help preserve ToM abilities in older adults, mitigating age-related declines.

In the current study, bilingualism is examined through individual variations in the participants' diverse language experiences rather than a binary comparison ToM between monolinguals and bilinguals. Recent perspectives on bilingualism have largely converged in agreement that bilingualism exists on a continuum rather than as a strict monolingual-bilingual dichotomy, and research has shifted away from traditional designs with monolingual versus bilingual comparisons and toward a model of bilingualism as a spectrum of experiences that may affect the structure and function of the brain (e.g., DeLuca et al., 2019; Luk & Bialystok, 2013; Yow & Li, 2015). Here, to quantify individuals' degree of bilingualism and its role in the trajectories of ToM aging, we first extract self-reported measures of bilingualism-related variables on the timing of second language learning (i.e., second language age of acquisition [L2AoA]), usage, and proficiency in the different languages, the relative balance or diversity in language usage, as well as language-switching practices.

We then apply data reduction techniques (i.e., principal component analysis [PCA]) to create a more manageable set of variables for analysis (see Gullifer & Titone, 2020, for a similar approach). By modeling bilingualism as a multifaceted experience, we aim to examine whether the various experiences of bilingualism contribute to better ToM skills across young and older adulthood.

When examining the effects of aging and bilingualism in relation to ToM, it is also important to consider other behavioral correlates, including general cognitive capacities such as processing speed, memory, and executive functions. Specifically, studies documenting age-related differences in ToM have produced contradictory findings regarding whether such differences are explained by decreased *ToM competence* or decreased *cognitive competence*, as general cognitive declines often co-occur with ToM declines in these studies (e.g., Charlton et al., 2009; Phillips et al., 2011; Sandoz et al., 2014; see Hamilton et al., 2022, for a review). Some studies reported that older adults' ToM deficits could be partially or fully explained by a decline in episodic memory (Fischer et al., 2017), executive functions (e.g., updating information in working memory, or inhibitory control; Phillips et al., 2011; Rahman et al., 2021), or processing speed (Charlton et al., 2009). Other studies provided evidence suggesting that age differences in ToM were independent of general cognitive decline such that age remained a significant predictor of ToM performance even after controlling for general cognition (e.g., Cavallini et al., 2013; Grainger et al., 2023; Kong et al., 2022). Compounding the issue further is the potential positive effect of bilingualism in maintaining general cognitive ability in older bilinguals (e.g., Ballarini et al., 2023; Chan et al., 2020; Schroeder & Marian, 2012; see Chen et al., 2022, or Ware et al., 2020, for meta-analyses). Although the argument that bilingualism confers a cognitive advantage in older adults has been challenged by recent replication and meta-analytic studies (e.g., Papageorgiou et al., 2019; Samuel et al., 2018), it is nevertheless important to control for such variables in our study.

In summary, our study assesses young and older adults on a battery of ToM tasks that evaluated ToM understanding across different types of content (e.g., inferences about others' beliefs, intentions, or emotions) and levels of complexity (e.g., inferences about basic or higher-order ToM). Our hypotheses are twofold: first, we expect a main effect of age, with older adults overall performing poorer than young adults on ToM tasks. Second, we predict that bilingualism will moderate ToM ability in older adults, due to its potential compensatory effect in supporting social cognitive function against age-related declines. It is possible that older adults with an earlier onset age of bilingualism and/or more extensive bilingual experience would outperform those with later acquisition and/or less extensive experience. However, we did not have a specific hypothesis regarding differential effects for L2AoA and other aspects of bilingualism as prior research on younger adults has yielded mixed findings about which components of bilingualism are associated with ToM (e.g., Navarro et al., 2022; Tiv et al., 2021). To determine whether age and bilingualism effects are independent of general cognition, we also included assessments of processing speed, episodic memory, working memory, and inhibitory control to control for the possible contribution of these variables in explaining participants' ToM performance in our study.

## 2. Methods

### 2.1. Participants

Our sample consisted of 185 participants, including 80 young adults (YA;  $M_{\text{age}} = 22.03$  years, range = 19–30; 47 females, 33 males)

and 105 older adults (OA;  $M_{age} = 66.23$  years, range = 56–79; 69 females, 36 males). An a priori power analysis suggested that a minimum sample size of 128 (64 per group) is required to achieve 80% power ( $\alpha = .05$ ) in detecting a moderate effect of age ( $d = 0.5$ ) based on previous meta-analyses (e.g., Henry et al., 2013). However, we recruited a larger sample size to examine the effects of bilingualism variables using regression analyses. We conducted simulation-based power analysis using the “simr” package (Green & MacLeod, 2016) in R to determine the achieved power to detect a significant main effect for the tested predictors in the Poisson regression model, which showed that our sample size was sufficient to achieve 80% power ( $\alpha = .05$ , 1000 simulations). The young adult sample was mainly undergraduate students and recruited through email or announcements in lectures. Older participants were community residing and recruited via posters in local care service centers, on social media, or through word-of-mouth. Participants received either course credit (students only) or SGD30 gift vouchers for their participation. The study protocol was approved by the Institutional Review Board of the Singapore University of Technology and Design (approval number 16–109 and 21–425). Participant demographic characteristics (means and SDs) and Welch’s  $t$ -test statistics, which are appropriate for group comparisons with unequal sample sizes (Delacre et al., 2017), are presented in Table 1. A detailed report of descriptive statistics and distribution for all variables in this study can be found in Supplementary Materials (Table S1).

All participants were Singapore citizens except for three young adults who were not Singapore citizens but reported having lived in Singapore for at least 10 years. All were ethnic Chinese. All participants reported having normal/corrected-to-normal visual acuity, normal color vision, and no history of neurological or psychiatric illnesses. Additionally, older adults were screened for abnormal cognitive decline using the Singapore version of the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) that was validated for use in the local older population (e.g., Ng et al., 2015). We used the recommended cutoff score of 23 (out of 30), with those scoring 22 or below suspected of having mild cognitive impairment; this resulted in the exclusion of three additional older adults (scores are 22, 19, and 18) who participated in the study but were not included in the sample reported here. Older participants scored, on average, 27.21 on the MoCA test ( $SD = 1.71$ , range = 23–30). Information about education (i.e., the highest level of education completed) was collected to index participants’ socioeconomic status. We coded education on a 5-point scale: 1 = *primary school*, 2 = *secondary school*, 3 = *high school, diploma, or junior college*, 4 = *undergraduate*, 5 = *post-graduate*. Older adults’ education ranged from 1 to 5 with a mean of 2.75, which was significantly lower compared with young adults ( $M = 3.24$ , range = 3–5),  $t_{Welch}(153.77) = 4.41$ ,  $p < .001$ ,  $d = .63$ . Hence, education was also included as a control variable in further analyses of the age effect. Given mixed evidence on gender differences in domains of ToM (Dodell-Feder et al., 2020; Fischer et al., 2017), we also checked for gender differences and gender by age interactions in ToM understanding, but none were found,  $ps > .83$ ; thus, gender was not considered in the further analyses.

All participants completed a questionnaire on their language background (see 2.2.2. Language background measures, for more details). All participants reported speaking two or more languages; 25 reported knowing two languages (20 YA and 2 OA knew English and Mandarin, 2 OA knew Mandarin and Cantonese, and 1 OA knew Mandarin and Hokkien), and 160 knew three or more languages (60 YA and 98 OA knew English, Mandarin, and one or two

**Table 1.** Descriptive statistics and analysis of age group differences for participants’ demographic, language, cognitive, and ToM measures

|   | Young adults<br>( $n = 80$ ) | Older adults<br>( $n = 105$ ) |             |
|---|------------------------------|-------------------------------|-------------|
|   | Mean (SD)                    | Mean (SD)                     | $t_{Welch}$ |
| Age (years)   | 22.03 (2.09)                 | 66.23 (5.66)                  | −73.64***   |
| Education (1–5)                                       | 3.24 (0.46)                  | 2.75 (1.00)                   | 4.41***     |
| MoCA (23–30)  | –                            | 27.21 (1.71)                  | –           |
| Language variables                                    |                              |                               |             |
| L2AoA (years)   | 0.76 (1.83)                  | 4.90 (2.80)                   | −12.14***   |
| Proficiency of most proficient language (1–10)        | 9.31 (0.99)                  | 9.30 (0.89)                   | 0.08        |
| Proficiency of second-most proficient language (1–10) | 7.24 (1.69)                  | 8.09 (1.64)                   | −3.42***    |
| Usage of most used language (0–1)                     | .78 (.15)                    | .71 (.17)                     | 2.93**      |
| Usage of second-most used language (0–1)              | .19 (.14)                    | .22 (.14)                     | −1.72       |
| Weighted language entropy (0–2) <sup>a</sup>          | 0.59 (.34)                   | 0.82 (.38)                    | −4.51***    |
| Single-language context                               | .61 (.33)                    | .50 (.37)                     | 2.12*       |
| Dual-language context                                 | .21 (.24)                    | .32 (.30)                     | −2.76**     |
| Dense-Codeswitching context                           | .18 (.21)                    | .18 (.24)                     | 0.01        |
| PC1: bilingual usage                                  | −0.27 (0.95)                 | 0.20 (0.99)                   | −3.29**     |
| PC2: bilingual proficiency                            | −0.06 (0.99)                 | 0.04 (1.01)                   | −0.69       |
| PC3: code-switching                                   | 0.15 (0.84)                  | −0.11 (1.10)                  | 1.86        |
| Cognition <sup>b</sup>                                |                              |                               |             |
| Processing speed                                      | 0.86 (0.59)                  | −0.66 (0.71)                  | 15.84***    |
| Episodic memory                                       | 0.34 (0.88)                  | −0.26 (1.01)                  | 4.29***     |
| Working memory  | 0.54 (0.67)                  | −0.41 (1.02)                  | 7.69***     |
| Inhibitory control                                    | 0.00 (0.90)                  | 0.00 (1.08)                   | −0.01       |
| ToM task performance                                  |                              |                               |             |
| Number of error responses                             | 0.55 (0.84)                  | 1.68 (1.45)                   | −6.63***    |

Note: ToM = theory of mind; MoCA = Montreal Cognitive Assessment; L2AoA = second language age of acquisition; PC = principal component. Education was coded on a 1–5 scale: 1 = *primary school*, 2 = *secondary school*, 3 = *high school, diploma, or junior college*, 4 = *undergraduate*, 5 = *post-graduate*. Language usage was self-reported frequency of usage in a typical week (in proportion), and proficiency was the average score for listening and speaking on a scale from 1 = *not proficient* to 10 = *very proficient*.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

<sup>a</sup>Scores range from 0 indicating usage of only one language across various social contexts (i.e., home, school, work, and others), to a maximum of 2 indicating equal usage of all different languages, hence, an integrated language context.

<sup>b</sup>Dependent measures for the cognitive assessments are expressed as z-scores, with higher scores indicative of better performance. The four domains of processing speed, episodic memory, working memory, and inhibitory control were assessed by the Digit Symbol Substitution Test, Rey Auditory Verbal Learning Test, 2-back task, and numeric Stroop task, respectively.

of the following languages: Chinese dialects such as Cantonese, Hokkien, Teochew, etc., and other languages such as Malay, Japanese, Korean, German, French, etc., while the remaining 2 OA reported knowing English, Malay, and Chinese dialects). However, these multilingual participants reported minimal usage of their third-most-used ( $M_{YA-usage3} = 4.95\%$ ,  $M_{OA-usage3} = 7.07\%$ ) and fourth-most-used ( $M_{YA-usage4} = 2.66\%$ ,  $M_{OA-usage4} = 2.45\%$ )

languages as compared with their two most used languages (see Table 1). Hence, we used the term “bilinguals” to refer to the participants in this study. For detailed reports of descriptive statistics for participants’ language variables and comparisons between age groups, refer to Table 1.

## 2.2. Materials and measurements

All tasks and questionnaires described in this section have validated versions in English and Chinese. During recruitment, participants indicated which language (English or Chinese) they were more dominant and most comfortable speaking in, and this information serves as the language of testing for the respective participants. All 80 young participants and 90 older participants completed the study in English, while 15 older participants completed the study in Chinese. For all tasks except the verbal memory test, the English and Chinese versions use the same stimuli and only differ in their instruction language. For the verbal memory test, different stimuli are used, but both versions have been locally validated (Lee et al., 2012).

### 2.2.1. Theory-of-mind assessments

The Theory of Mind Task Battery (ToMTB; Hutchins et al., 2008) was used to examine participants’ ToM abilities. Although the ToMTB was initially developed for use in children (as is the case with most ToM tasks), it has been used to assess ToM in older adults with and without neurocognitive disorders with good validity (e.g., De Rezende et al., 2018; Ferreira Pereira et al., 2022). The task battery consists of a total of 15 questions across nine tasks that assess ToM understanding across a variety of content and complexity: (a) basic emotion recognition task, (b) desire-based emotion task, (c) perception-based belief task, (d) visual perspective-taking task, (e) perception-based action task, (f) first-order false belief task, (g) belief- and reality-based emotion and second-order emotion task, (h) message-desire discrepant task, and (i) second-order false belief task (see Table S2 in Supplementary Materials, for more details). There were four questions for the task (a), two questions for the task (d), three questions for the task (g), and one question for each of the other tasks.

Tasks were presented as short vignettes that appeared in a story-book format. Participants listened to the experimenter who narrated the vignettes, along with colored visual illustrations. For each vignette, participants were to answer a question about the main character’s beliefs, intentions, or emotions. The experimenter introduced the test by stating, “I am going to read you some short stories and ask you questions about the story. You can answer with words, or you can point to the answer.” For each question, there is one correct response option and three plausible distractors, where any incorrect answer is deemed an error response (examples can be found at <https://bit.ly/task-battery>). We calculated the number of error responses across all test questions as an index of ToM ability. ToM scores range from 0 to 15, with lower scores indicating fewer ToM errors and hence better ToM understanding.

### 2.2.2. Language background measures

We used a questionnaire adapted from the Language Background Questionnaire (LBQ; Yow & Li, 2015) and the Bilingual Interactional Context Questionnaire (BICQ; Hartanto & Yang, 2019) to assess participants’ language background. We constructed the questionnaire in such a way that it captures different aspects of a bilingual experience within a multilingual environment like Singapore. While items from the LBQ mainly assess bilinguals’ general language experiences, such as the age of acquisition, usage, proficiency, and

balance (see Dass et al., 2024), items from the BICQ examine bilinguals’ diverse code-switching and interactional contexts, a “notable characteristics of the Singapore linguistic environment” (Ooi et al., 2018, p. 869). Participants reported the age at which they were first exposed to each of the languages they knew. Language proficiency in listening and speaking was reported on a 10-point scale (1 = *not proficient* to 10 = *very proficient*) and then averaged to obtain one proficiency score per language. Participants were also asked about the amount of time (in percent) spent in different social contexts (i.e., home, work, school, and others) in a typical week, how much they used each language (in proportion) in each of these contexts, and how often (in proportion) they engaged in different interactional contexts (i.e., single-language, dual-language, and dense code-switching; see Green & Abutalebi, 2013). To quantify the diversity of language use in participants’ environments, we computed a weighted language entropy score across the different social contexts per participant (following Gullifer & Titone, 2018; Li et al., 2021). Weighted language entropy scores range from 0 indicating usage of only one language across contexts, to 2 indicating equal usage of all possible languages (i.e., a completely integrated context). The group means and SDs for the language variables are reported in Table 1, with a more detailed report of descriptive statistics and correlations presented in Supplementary Materials (Tables S1 and S3).

While we treat bilingualism as a multifaceted experience with multiple language variables, we would like to identify key patterns of the multivariate data for meaningful analyses and interpretation. Thus, we conducted a PCA to analyze and reduce the dimensionality of these language variables. By construction, all principal components extracted from a PCA are orthogonal (i.e., uncorrelated) to each other (Kassambara, 2017). The variable L2AoA was not included in the PCA as it did not correlate with any language variables ( $r_s < .20$ ,  $p_s > .22$ ; see Supplementary Table S3). Data were standardized before performing PCA. Using the “princomp()” function of the “stats” R package, three principal components were extracted (determined by eigenvalues  $> 1$ ; also see scree plot in Supplementary Figure S1), which explained 79.1% of the total variance. Table 2 shows the component loadings, which indicate the contributions of variables to each component. Inspection of the variables loaded on each component suggests that Principal

**Table 2.** Loading matrix of the three principal components for bilingual language experience

|   | Component 1 | Component 2 | Component 3 |
|---|-------------|-------------|-------------|
| Proficiency of most proficient language     | .05         | <b>.79</b>  | .06         |
| Proficiency of 2nd-most proficient language | .26         | <b>.56</b>  | -.05        |
| Usage of most used language                 | <b>-.45</b> | .14         | -.11        |
| Usage of second-most used language          | <b>.42</b>  | -.17        | .13         |
| Weighted language entropy                   | <b>.42</b>  | -.03        | -.01        |
| Single-language context                     | <b>-.43</b> | .04         | .12         |
| Dual-language context                       | .32         | -.04        | <b>-.71</b> |
| Dense-codeswitching context                 | .28         | -.02        | <b>.67</b>  |

Note:  $N = 185$ . Loadings greater than 0.4 (bolded) are considered stable (Stevens, 2009).

Component 1 represents “bilingual usage,” Principal Component 2 represents “bilingual proficiency,” and Principal Component 3 represents “code-switching.” More specifically, the “bilingual usage” component had relatively high positive loadings ( $> .40$ ) for the usage of the second-most used language and weighted language entropy, and high negative loadings for the usage of the most used language and single-language context (explaining 49.9% of the total variance). The “bilingual proficiency” component had high positive loadings for both proficiency values of the two most proficient languages (16.1% of total variance). The “code-switching” component had a negative loading for dual-language context and a positive loading for dense code-switching context (13.3% of total variance). Next, component scores were extracted via “get\_pca\_ind()” in R, which computed standardized component scores for each participant and each component by multiplying the individual’s variable-level responses by the component loadings. Based on the attributes of the three principal components, higher component scores for PC1, PC2, and PC3 indicate more balanced bilingual usage, higher proficiency in their two languages, and more dense code-switching interactional experience, respectively. The three component scores, together with L2AoA, are the main measures of bilingualism in this study and are used in further analysis to examine the effects of bilingualism on ToM.

### 2.2.3. General cognition assessments

We used a battery of tasks to assess four general cognitive domains: processing speed, episodic memory, working memory, and inhibitory control (see [Supplementary Materials](#) for details). The Digit Symbol Substitution Test (DSST) of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 2000) was used to index the speed of processing. The Rey Auditory Verbal Learning Test (RAVLT; Schmidt, 1996) was administered, and the delayed recall score from this task was used as a measure of verbal episodic memory. We computerized the 2-back task and the numeric Stroop task to assess working memory and inhibition, respectively. Raw scores for each of the individual cognitive assessments were transformed into  $z$ -scores based on the sample means and  $SD$ s across all participants, with higher scores indicating better performance.

### 2.3. Procedure

Participants were tested individually in a quiet room after providing informed consent. All older participants were assessed on the MoCA at the beginning of the experiment session. All young and older participants completed the assessments in the following order: Stroop, RAVLT, DSST, ToMTB, 2-back, and language background measures.

### 2.4. Analysis

The preliminary analyses revealed no missing data and no multivariate outliers. As a result, all analyses reported in this article included the full sample ( $N = 185$ ), unless otherwise noted. Skewness and kurtosis values for all variables were within the acceptable range of  $(-3, +3)$  and  $(-10, +10)$ , respectively (see [Supplementary Tables S1 and S4](#)), indicating no violation of the normality assumption (Kline, 2011).

To examine the effects of age and bilingualism on ToM scores (i.e., number of errors), we conducted a series of generalized linear models (GLM) with Poisson distribution<sup>1</sup>, which gives the

<sup>1</sup>We also tried to fit the data with a simpler linear regression model, which showed that the assumption of homoscedasticity (uniform distribution of residuals) was not met.

probability of an event happening a certain number of times and is suitable to model count data (Zeileis et al., 2008). Since Poisson regression models the log of the expected count as a linear function of predictor variables (e.g.,  $\log(\mu) = \alpha + \beta x$ ), the regression coefficient can be interpreted as follows: for a unit change in the predictor variable, the estimated count is expected to change by a factor of the exponentiated coefficient (i.e.,  $e^\beta$ ). This factor is referred to as a *rate ratio*, representing a percent change in the response for a unit change in the predictor variable. To estimate model parameters, ToM scores were fitted with hypothesized predictors and covariates using the “glm()” function of the “stats” R package. Model comparisons were done via a chi-squared test using the “anova()” function. We checked the model assumptions using the “DHARMA” package. For all models reported below, no evidence of overdispersion or zero-inflation was detected, and residuals followed uniform distributions with no heteroscedasticity. VIFs for all predictors were  $< 3$ , indicating no multicollinearity concern (Zuur et al., 2010; see [Table S5 in Supplementary Materials](#)). For results of the final models, we report beta coefficients (i.e., estimates of standardized predictors),  $z$  tests,  $p$  values, and rate ratios in the main text, with more detailed reports of model estimates, standard errors, 95% confidence intervals, and model fit statistics (i.e., chi-squared tests, AIC, and Nagelkerke/Cragg-Uhler’s pseudo  $R^2$ ) shown in [Table 3](#).

The first set of GLM analyses investigated whether there is an age-related decline in ToM performance. As we were interested in whether there are age differences between young and older groups, we constructed two nested models: Model 1 only included control variables of education, processing speed, episodic memory, working memory, and inhibitory control, and Model 2 included age group (dummy coded; young adults = 0, older adults = 1) as an additional predictor in the model. We predicted that age group is a significant predictor of the expected number of errors in the ToM tasks, such that overall, older adults performed poorer (i.e., more errors) than young adults.

Next, we examined whether bilingualism measures (i.e., L2AoA and three component scores derived from the PCA) explain significant variance in ToM performance beyond the age effects observed in our data. To test whether each bilingualism measure independently contributes to variance in ToM performance, we constructed four separate models, each including one bilingualism measure (i.e., L2AoA, PC1 “bilingual usage,” PC2 “bilingual proficiency,” and PC3 “code-switching”) as a predictor, as well as age group and all control variables. Since our hypothesis was to examine the effects of various aspects of the bilingual experience on ToM abilities in aging, we also constructed a full model (Model 3) that included all four bilingualism measures as predictors, together with the predictor of age group and control variables in the same model.

Finally, we examined additional models treating age as a continuous variable to explore how ToM performance changes as a function of increasing age. Because our sample did not include all ages within the full range (with a gap in ages between 30 and 56), we conducted separate analyses for young adults and older adults. We ran similar GLM analyses as in Models 1–3 and treated age as a continuous variable in these models. Model comparisons and assumption checks were the same as those conducted in Models 1–3.

<sup>2</sup>Nagelkerke/Cragg-Uhler’s pseudo  $R^2$  provides information similar to that provided by  $R^2$  in ordinary least squares regression, but it is computed based on the log likelihoods of a Poisson model (Nagelkerke, 1991). The pseudo  $R^2$  ranges from 0 to 1, with 1 indicating that the full model perfectly predicts the outcome.

**Table 3.** Results from generalized linear models predicting theory-of-mind task performance (number of error responses)

| Predictors            | Model 1                              |           |                |                    | Model 2                              |           |                |         | Model 3                              |           |                |                   |
|-----------------------|--------------------------------------|-----------|----------------|--------------------|--------------------------------------|-----------|----------------|---------|--------------------------------------|-----------|----------------|-------------------|
|                       | <i>b</i>                             | <i>SE</i> | 95% CI         | $\beta$            | <i>b</i>                             | <i>SE</i> | 95% CI         | $\beta$ | <i>b</i>                             | <i>SE</i> | 95% CI         | $\beta$           |
| (Intercept)           | 0.33                                 | 0.28      | [-0.22, 0.89]  |                    | -0.08                                | 0.31      | [-0.69, 0.51]  |         | -0.26                                | 0.31      | [-0.87, 0.35]  |                   |
| Education             | -0.10                                | 0.09      | [-0.28, 0.08]  | -0.08              | -0.12                                | 0.09      | [-0.30, 0.06]  | -0.10   | -0.08                                | 0.09      | [-0.26, 0.09]  | -0.07             |
| Processing speed      | -0.23                                | 0.08      | [-0.39, -0.06] | -0.23**            | -0.01                                | 0.11      | [-0.22, 0.20]  | -0.01   | -0.01                                | 0.11      | [-0.22, 0.21]  | -0.01             |
| Episodic memory       | -0.13                                | 0.07      | [-0.27, 0.00]  | -0.13 <sup>+</sup> | -0.14                                | 0.07      | [-0.28, -0.00] | -0.14*  | -0.14                                | 0.07      | [-0.28, -0.00] | -0.14*            |
| Working memory        | -0.21                                | 0.07      | [-0.36, -0.06] | -0.21**            | -0.16                                | 0.08      | [-0.31, -0.01] | -0.16*  | -0.16                                | 0.08      | [-0.31, -0.01] | -0.16*            |
| Inhibitory control    | 0.02                                 | 0.06      | [-0.11, 0.14]  | 0.02               | -0.002                               | 0.06      | [-0.13, 0.12]  | -0.002  | 0.00                                 | 0.06      | [-0.12, 0.13]  | -0.002            |
| Age group: Older      |                                      |           |                |                    | 0.77                                 | 0.23      | [0.31, 1.23]   | 0.77**  | 0.51                                 | 0.26      | [0.00, 1.03]   | 0.51 <sup>+</sup> |
| L2AoA                 |                                      |           |                |                    |                                      |           |                |         | 0.06                                 | 0.03      | [0.01, 0.12]   | 0.20*             |
| Bilingual usage       |                                      |           |                |                    |                                      |           |                |         | 0.03                                 | 0.07      | [-0.11, 0.17]  | 0.03              |
| Bilingual proficiency |                                      |           |                |                    |                                      |           |                |         | 0.07                                 | 0.07      | [-0.08, 0.21]  | 0.07              |
| Code-switching        |                                      |           |                |                    |                                      |           |                |         | 0.06                                 | 0.07      | [-0.07, 0.19]  | 0.06              |
| Model fit             | $\chi^2(5) = 63.01, p < .001$        |           |                |                    | $\chi^2(6) = 74.11, p < .001$        |           |                |         | $\chi^2(10) = 80.81, p < .001$       |           |                |                   |
|                       | AIC = 524.14                         |           |                |                    | AIC = 515.04                         |           |                |         | AIC = 516.34                         |           |                |                   |
|                       | Pseudo- <i>R</i> <sup>2</sup> = .302 |           |                |                    | Pseudo- <i>R</i> <sup>2</sup> = .345 |           |                |         | Pseudo- <i>R</i> <sup>2</sup> = .370 |           |                |                   |

Note: *N* = 185. Data were fitted to Poisson distribution with a log link. L2AoA = second language age of acquisition; AIC = Akaike information criterion. No multicollinearity was detected in the regression models. Pseudo-*R*<sup>2</sup> was computed based on log-likelihoods for each model, indicating how well the model explains the data (Nagelkerke, 1991). <sup>+</sup>*p* < .06. \**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

### 3. Results

#### 3.1. Age differences in ToM

As shown in Table 3, Model 1 (with control variables only) was significant,  $\chi^2 = 63.01, N = 185, df = 5, p < .001$ . Processing speed (DSST score;  $\beta = -0.23, z = -2.72, p = .007, \text{rate ratio} = 0.80$ ) and working memory (accuracy in 2-back task;  $\beta = -0.21, z = -2.83, p = .005, \text{rate ratio} = 0.81$ ) were significant predictors of performance in ToM. The effect of episodic memory was marginally significant (delayed verbal recall from RAVLT;  $\beta = -0.13, z = -1.95, p = .051, \text{rate ratio} = 0.87$ ). The other control variables (education and inhibitory control) were not significant, *ps* > .29. Overall, faster processing speed and better memory were associated with fewer errors in ToM understanding.

Model 2 which included age group as a predictor variable provided an overall good fit to the data,  $\chi^2 = 74.11, N = 185, df = 6, p < .001$ . It also yielded a significantly better fit than Model 1,  $\chi^2 = 11.10, N = 185, df = 1, p < .001$ . Results revealed a significant effect for age group,  $\beta = .77, z = 3.28, p = .001, \text{rate ratio} = 2.15$ , indicating that the number of ToM errors made by older adults was 2.15 times the number of errors made by young adults while holding the control variables (i.e., education and general cognition) constant.<sup>3</sup> Note that the two control variables of episodic memory ( $\beta = -0.14, z = -2.01, p = .044, \text{rate ratio} = 0.87$ ) and working memory ( $\beta = -0.16, z = -2.15, p = .032, \text{rate ratio} = 0.85$ ) individually contributed to the performance in ToM, suggesting that better episodic and working memory led to less errors and hence better performance in ToM tasks. The other control variables (i.e., education, processing speed, and inhibitory control)

<sup>3</sup>A generalized linear model with age group as the single predictor, without any control variables, confirmed that the simple effect of age group was significant,  $\beta = 1.11, SE = 0.17, 95\% \text{ CI } [0.78, 1.44], z = 6.61, p < .001, \text{rate ratio} = 3.05$ . This single-predictor model was significant:  $\chi^2 = 52.97, df = 1, p < .001$ .

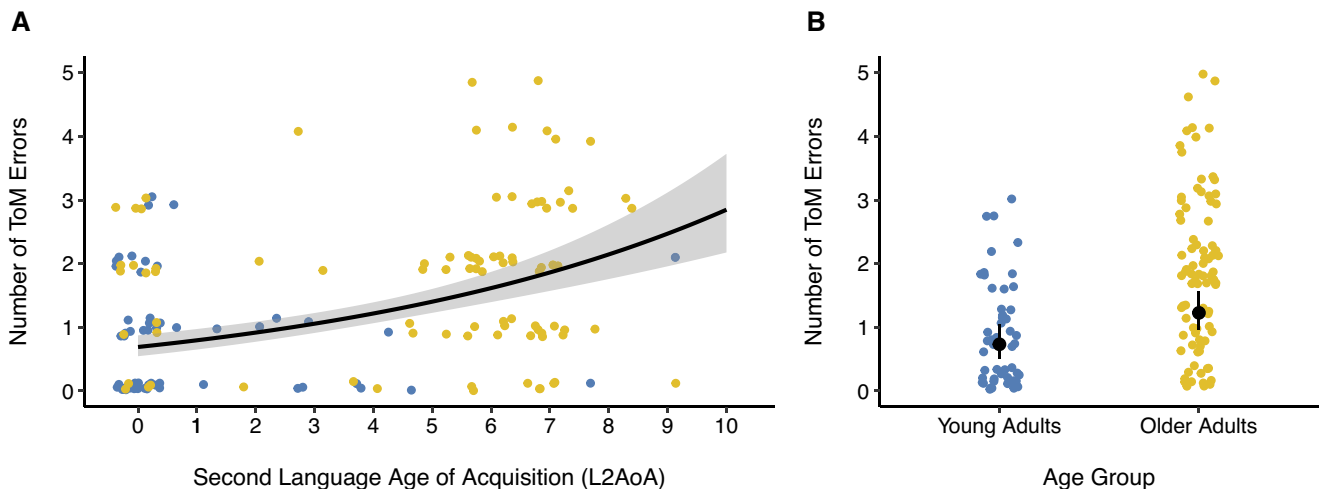
were not significant, *ps* > .18. These results provided evidence for an age-related decline in ToM abilities in older adults compared with young adults, which was independent of age-based differences in education experience and general cognition.

#### 3.2. Effects of bilingualism on ToM

Model comparisons that included one bilingualism measure (i.e., L2AoA, PC1, PC2, or PC3) as a predictor in addition to age group and control variables found that only the model with L2AoA (and without any PC) yielded a significantly better fit compared with the model without this predictor (i.e., Model 2), *N* = 185,  $\chi^2 = 4.86, df = 1, p = .027$ . In contrast, models with an additional predictor of PC1, PC2, or PC3 did not yield a significantly better fit than Model 2 (all *ps* > .39), indicating no significant contribution of the three component scores to variance in ToM.

Model 3, which included L2AoA, PC1, PC2, PC3, age group, as well as control variables, provided a good fit to the data,  $\chi^2 = 80.81, N = 185, df = 10, p < .001$  (see Table 3). Additionally, we evaluated the two-way interaction effects between each bilingualism variable and age group by examining whether the inclusion of the interaction term provided a better fit than the model without the interaction term, but none were significant, all *ps* > .23. The nonsignificant interactions were further confirmed by additional post hoc power analyses showing that it requires either a substantial effect size (rate ratio larger than 2.39) or a sample size of at least 5000 to detect a significant interaction effect with 80% power.

In Model 3, L2AoA emerged as a significant predictor of performance in ToM,  $\beta = 0.20, z = 2.22, p = .027, \text{rate ratio} = 1.06$ , where 1 year older in the onset age of bilingualism increased the number of ToM errors by 6% (see Figure 1A). This suggested that earlier bilingual acquisition was associated with better ToM performance within the current combined sample that included both young and older ages. The effect for the age group only reached



**Figure 1.** Effects of L2AoA and age group predicting theory-of-mind task performance.  $N = 185$ . Each colored dot represents an individual participant (jittered). The line in graph (A) represents the estimated regression function between ToM scores and L2AoA obtained from Model 3, holding other variables constant. The black dots in graph (B) represent the estimated group means for young and older adults from Model 3, controlling for other variables, with error bars indicate 95% confidence intervals. There is no significant interaction effect between age group and L2AoA on the number of ToM errors.

marginal significance in Model 3,  $\beta = -0.51$ ,  $z = 1.96$ ,  $p = .050$ , rate ratio = 1.67 (see Figure 1B). Controlling for the effects of various bilingual language experiences seemed to attenuate the age effects on ToM. However, unlike the experience of early or late L2AoA, the other three aspects of bilingual experience were not significant predictors of ToM: bilingual usage,  $\beta = 0.03$ ,  $z = 0.47$ ,  $p = .64$ , rate ratio = 1.03; bilingual proficiency,  $\beta = 0.07$ ,  $z = 0.94$ ,  $p = .35$ , rate ratio = 1.07; and code-switching,  $\beta = 0.06$ ,  $z = 0.90$ ,  $p = .37$ , rate ratio = 1.06. Meanwhile, episodic memory ( $\beta = -0.14$ ,  $z = -1.96$ ,  $p = .0498$ , rate ratio = 0.87) and working memory ( $\beta = -0.16$ ,  $z = -2.05$ ,  $p = .041$ , rate ratio = 0.85) explained significant variance in ToM performance, where better episodic and working memory was associated with better ToM. These results revealed that individuals' early bilingual experience (in particular, L2AoA) contributed to performance in ToM over and beyond the impact of age, education, and cognitive aging on ToM.

### 3.3. Additional analyses – age as a continuous variable

We conducted additional GLMs treating age as a continuous variable for young adults and older adults separately due to a gap in our samples of those ages between 30 and 56. For the young adult group ( $N = 80$ ), although the full model—including all predictors (age, L2AoA, bilingual usage, bilingual proficiency, and code-switching) and control variables—demonstrated a good overall fit,  $\chi^2 = 19.47$ ,  $N = 80$ ,  $df = 10$ ,  $p = .035$ , none of the predictors significantly explained the interindividual variance in ToM, all  $ps > .25$ . For the older adult group ( $N = 105$ ), the same full model also yielded a good overall fit,  $\chi^2 = 33.20$ ,  $N = 105$ ,  $df = 10$ ,  $p < .001$ . Importantly, both age and L2AoA emerged as significant predictors of ToM performance in older participants (age:  $\beta = 0.27$ ,  $z = 2.83$ ,  $p = .005$ , rate ratio = 1.32; L2AoA:  $\beta = 0.18$ ,  $z = 2.18$ ,  $p = .029$ , rate ratio = 1.20). Specifically, the number of ToM errors was expected to increase as age and L2AoA increase in older adults, controlling for other aspects of bilingual experience, education, and cognitive ability. None of the other bilingualism measures and control variables contributed to the variance of ToM performance in older adults, all  $ps > .12$ .

## 4. Discussion

The present research examined the effects of age and bilingualism on the ToM performance of young and older adults. Overall, we found that older adults committed more errors in ToM tasks than young adults. Most importantly, we found that an earlier L2AoA predicted better ToM performance in our sample of young and older bilinguals, over and above the effects of aging. We did not find significant effects of other aspects of bilingual experience, such as bilingual usage, bilingual proficiency, and code-switching interactional experience on ToM performance. The findings suggest that bilingualism, particularly early second language acquisition, appears to contribute to intact ToM performance against normal age-related deteriorations.

In line with our hypothesis, results revealed an overall negative effect of age on ToM performance and this effect of age was robust even when other individual cognitive factors such as education and general cognition were controlled for. Our finding that cognitively healthy older adults (aged 56–79) performed significantly worse than young adults (aged 18–30) in ToM tasks concurs with current converging evidence that older adults face increased difficulties in mental-state reasoning across various ToM tasks, including perspective taking (e.g., Bradford et al., 2023), false-belief understanding (e.g., Phillips et al., 2011), and/or inference of complex emotions and beliefs (e.g., Greenberg et al., 2023; Krendl et al., 2023; cf. Dodell-Feder et al., 2020). Moreover, we found that the age difference in participants' ToM remained significant even after controlling for individual variations in education, processing speed, episodic memory, working memory, and inhibitory control, even though better episodic memory and better working memory were also found to be associated with better ToM performance. This finding is congruent with previous studies suggesting a specific ToM impairment in old age that is independent of educational experience and age-related changes in general cognition during healthy aging (e.g., Cavallini et al., 2013; Kong et al., 2022; cf. Johansson Nalaker et al., 2018; Murphy et al., 2019).

A significant finding from the current research is that early bilingualism, operationalized as an earlier L2AoA, has a positive effect on ToM performance, specifically within the group of older



adults. In support of our hypothesis regarding the role of bilingualism in ToM development, we found that earlier L2AoA predicted better ToM performance across both groups of young and older participants, over and beyond the effect of age. Additional analyses revealed that the benefits of an earlier L2AoA were significant and unique to older adults, who tend to experience increased difficulty maintaining ToM abilities as they age. Importantly, we found that L2AoA independently contributed to ToM performance when other experience factors such as education and general cognitive abilities were controlled for, suggesting that earlier bilingual acquisition could directly enhance ToM abilities. This highlights that life-course factors, such as early bilingual experience, could offer opportunities to establish cognitive reserve related to social cognitive processes, as manifested in the performance of ToM tasks in our study. This corroborates past studies that suggested bilinguals' superior ToM performance in childhood appears to extend into adulthood (e.g., Navarro & Conway, 2021). In line with the STAC-r framework, lifelong (early onset) bilingual experience could act as an enriching factor that enhances brain structure and function relating to mental state reasoning (e.g., Li et al., 2024), thereby mitigating age-related declines.

We postulate that early bilingual experience contributes to the consolidation of social cognitive abilities in bilinguals. Previous evidence shows bilingual children's enhanced ability to infer others' intentions compared to monolingual children (e.g., Yow et al., 2017). While L2AoA may be considered as an index of prolonged bilingual language use (e.g., DeLuca et al., 2020), we argue that the effects of L2AoA observed in our study reflect the importance of early bilingual exposure rather than the sheer duration or frequency of bilingual language use. This is because, firstly, all young and older participants in our study reported being lifelong, active bilinguals, and secondly, our data showed no significant effect of individuals' degree of bilingualism in terms of usage and proficiency in different languages, and code-switching experiences. Early L2AoA may help preserve ToM abilities by enhancing neural plasticity early in life. For example, individuals with early L2AoA have shown greater functional connectivity between the left and right inferior frontal gyrus (IFG) than those with later L2AoA, indicating a greater ability among early bilinguals to manage their languages (Berken et al., 2016). This, in turn, may transfer to an understanding that others can have different mental states in relation to the same event from the self (i.e., "metalinguistic awareness" account; see Goetz, 2003). The advantage that early L2AoA has on ToM may persist into adulthood and allow early bilinguals to maintain these abilities against natural age-related declines. Nevertheless, given that our study was cross-sectional in nature, we are unable to assess any direct relationship between early bilingual experience and the developmental course of ToM in later adulthood in the absence of longitudinal data.

This finding is significant but should be considered within our model's constraints. First, the observed predictive effect of L2AoA ( $\beta = .20$ , rate ratio = 1.06) may be considered small (Cohen, 1988). The rate ratio 1.06 illustrated that 1 year difference in L2AoA was associated with a 6% difference in ToM scores. It is noted that studies on social processes or individual differences tend to produce smaller effects than experimental studies (e.g., Schäfer & Schwarz, 2019), likely due to larger samples in these former types of studies that have sufficient statistical power to detect small but significant effects. Second, our participants' bilingual language experience was assessed using self-reported measures including language proficiency. Past research suggests that there is a lack of correlation between self-ratings and objective proficiency measures (e.g.,

Tomoschuk et al., 2019), raising questions about the reliability of subjective measures in assessing participants' relative language proficiency. However, it is not always feasible to use objective measures to measure bilinguals' language proficiency. For example, the Multilingual Naming Test (MINT; Gollan et al., 2012) has been shown a reliable proficiency measure for bilinguals speaking English, Spanish, Mandarin, and Hebrew. However, there is no validated MINT or equivalent test in the other languages (e.g., Cantonese, Hokkien, Teochew, etc.) reported in our sample that could be used to assess participants' proficiency in these different languages. Future research could consider other ways to obtain objective measures of bilingual proficiency, such as developing a Cantonese, Hokkien, or Teochew version of a naming test, to further examine the effects of language proficiency on ToM in aging. Third, we found that better episodic and working memory were associated with higher performance in ToM. Consistent with previous studies (e.g., Jarvis & Miller, 2017; Phillips et al., 2011), ToM reasoning is dependent to some extent on these memory processes, as it involves remembering details and feelings associated with the event, as well as updating information about others' mental states when a situation changes. Nevertheless, the effects of early bilingual experience reported in the current study appear to be robust and remain significant beyond the positive influence of these general cognitive abilities on ToM across both younger and older adults.

One of the limitations of our study is that we considered ToM as a unified construct and used a composite score to represent participants' ToM ability. The ToM battery used in the present research encompasses a range of classic ToM tasks – emotional recognition, diverse desire, knowledge access, perspective taking, false belief, and complex emotion and belief inferences. However, as there are only a few questions that tap into each ToM component, it is not possible to evaluate the relationship between age-related changes and bilingualism in a specific aspect of ToM. It is possible that aging and bilingual experience are selectively associated with compromised performance on certain ToM aspects but not others (see Grainger et al., 2023, with evidence for "multidirectional" changes in four ToM-based tasks with age). For example, although meta-analytic work suggested that normal aging is likely associated with a decline in both cognitive and affective ToM (Roheger et al., 2022), there is evidence that cognitive ToM may be more affected by aging than affective ToM (Wang & Su, 2013). In fact, some researchers have suggested that ToM should not be treated as a dichotomous distinction between cognitive and affective components; instead, one should consider the possibility of multiple distinct or even overlapping processes within ToM (e.g., Navarro, 2022; Schaafsma et al., 2015). Therefore, future work could administer multiple representative tasks that target clearly defined processes of ToM (e.g., Fischer et al., 2017; Krendl et al., 2023) to understand the specificity of ToM changes as a function of age and diverse bilingual experience.

A key feature of the ToM assessment in our study, like many other ToM tasks in the literature, is the use of story vignettes and forced-choice response approach. Participants viewed static stimuli (vs. dynamic, naturalistic stimuli), and their answers were coded as correct or incorrect (i.e., binary). This limits the investigation of "real-world" mental state representation in adults with sufficient variance and sometimes may lead to ceiling effects in neurotypical adults (e.g., Yeung et al., 2024). In addition, while our current ToM measure included tasks with varying complexity (Hutchins et al., 2008), they may still be relatively simple compared with more complex ToM tasks, such as the Faux Pas or RMET. To better

capture individual differences in ToM performance among healthy young and older adults, recent trends suggest extending classical ToM paradigms with more dynamic and naturalistic stimuli (e.g., the Dynamic Theory of Mind Task, which uses clips from the sitcom *The Office*, see Krendl et al., 2023) or incorporating other measures such as open questions (e.g., the Edinburgh Social Cognition Test, see Baksh et al., 2018), reaction time (e.g., De Lillo & Ferguson, 2023), and eye tracking (e.g., Bradford et al., 2023). For example, Breil & Böckler (2020) recorded participants' eye movements while they performed a social video task that assessed empathy and ToM. Their results showed a substantial inter-individual variance in the time spent looking at the narrator during the videos, which was associated with ToM task performance. Future research could consider adopting these promising paradigms that allow for the investigation of more nuanced mental state representation in young and older adults. Another limitation of our study is that we used a fixed-order design, where all participants completed the ToM task battery after three cognitive assessments (i.e., Stroop, RAVLT, and DSST), which were then followed by a 2-back task and the language questionnaire. One reservation is whether this fixed-order design may have affected the ToM performance of older adults more than in young adults, as aging is associated with limited cognitive resources (e.g., Stern, 2002). Although we did not counterbalance the order of tasks, we ensured that participants took sufficient breaks when needed. Neither young nor older adults reported difficulty in understanding or engaging in the tasks, even for tasks that took place at the end of the session. Nevertheless, while it is unlikely that the significant age effects observed in our results were due to order effects, future research should consider counterbalancing the order of the tasks to control for such potential order effects.

Recent concerns has questioned the arbitrary classification of individuals as "older adults" versus "younger adults" in aging studies and the potential loss of information when age is treated as a categorical variable (Raimo et al., 2022). However, our study did not include participants aged between 30 and 56, which limits our ability to analyze age as a continuous variable across the entire sample. Thus, we grouped participants into young and older adult categories to facilitate direct comparisons of ToM performance and to examine the impact of bilingualism across distinct developmental periods. This approach is common in aging research to identify critical age-related differences, particularly when exploring the compensatory effects of bilingualism. However, such categorical groupings may fail to capture age-related changes in ToM that occur throughout the lifespan, particularly in middle-aged adults (see Bradford et al., 2023). Future research should consider alternative approaches, such as using a continuous age sample of participants (e.g., Grainger et al., 2023) to provide a more nuanced understanding of how aging and bilingual experience influence ToM across the lifespan.

In conclusion, the ability to maintain intact social cognitive function is important for social interactions with others in healthy aging. Life-course factors can provide neural resource enrichment and compensatory scaffolding that influence the trajectories of aging. Our study suggests that older adults who perform well in understanding the mental states of others might have benefited from compensatory mechanisms developed early in life. By examining the developmental time course of such influences, particularly the early onset of bilingualism, our study provides important insights into the cognitive reserve mechanism at different stages of the life course.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S1366728925000240>.

**Data availability statement.** The data and analytical code for the study are available on OSF at <https://doi.org/10.17605/osf.io/c69js>. The study reported in the manuscript was not preregistered.

**Funding statement.** We would like to express our appreciation to the participants who took part in the study. We thank Jia Wen Lee, Nina N. Ye, Juan Helen Zhou, Kwun Kei Ng, Joey Ju Yu Wong, and Janice Jue Xin Koi for their support during study design and data collection. This work was supported by the Ministry of Education Academic Research Fund Tier 2 (T2MOE2005) and the SUTD Kick-Start Initiative Grant (SKI 2021\_03\_11) to W. Quin Yow.

**Competing interest.** The authors declare no conflicts of interest.

## References

- Abutalebi, J., Canini, M., Della Rosa, P. A., Sheung, L. P., Green, D. W., & Weekes, B. S. (2014). Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiology of Aging*, 35(9), 2126–2133.
- Anderson, J. A., Grundy, J. G., Grady, C. L., Craik, F. I., & Bialystok, E. (2021). Bilingualism contributes to reserve and working memory efficiency: Evidence from structural and functional neuroimaging. *Neuropsychologia*, 163, 108071.
- Anderson, J. A., Hawrylewicz, K., & Grundy, J. G. (2020). Does bilingualism protect against dementia? A meta-analysis. *Psychonomic Bulletin & Review*, 27, 952–965.
- Baksh, R. A., Abrahams, S., Auyeung, B., & MacPherson, S. E. (2018). The Edinburgh Social Cognition Test (ESCoT): Examining the effects of age on a new measure of theory of mind and social norm understanding. *PLOS ONE*, 13(4), e0195818. <https://doi.org/10.1371/journal.pone.0195818>
- Ballarini, T., Kuhn, E., Röske, S., Altenstein, S., Bartels, C., Buchholz, F., ... & DELCODE study group (2023). Linking early-life bilingualism and cognitive advantage in older adulthood. *Neurobiology of Aging*, 124, 18–28.
- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. (1993). *Understanding other minds: Perspectives from autism*. Oxford University Press.
- Berken, J. A., Chai, X., Chen, J. K., Gracco, V. L., & Klein, D. (2016). Effects of early and late bilingualism on resting-state functional connectivity. *The Journal of Neuroscience*, 36(4), 1165–1172. <https://doi.org/10.1523/JNEUROSCI.1960-15.2016>
- Bialystok, E. (2021). Bilingualism: Pathway to cognitive reserve. *Trends in Cognitive Sciences*, 25(5), 355–364. <https://doi.org/10.1016/j.tics.2021.02.003>
- Bradford, E. E. F., Brunson, V. E. A., & Ferguson, H. J. (2023). Cognitive mechanisms of perspective-taking across adulthood: An eye-tracking study using the director task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(6), 959–973.
- Breil, C., & Böckler, A. (2020). The lens shapes the view: On task dependency in ToM research. *Current Behavioral Neuroscience Reports*, 7, 41–50.
- Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology*, 48, 257–270. <https://doi.org/10.1037/a0025402>
- Cavallini, E., Lecce, S., Bottiroli, S., Palladino, P., & Pagnin, A. (2013). Beyond false belief: Theory of mind in young, young-old, and old-old adults. *The International Journal of Aging and Human Development*, 76(3), 181–198.
- Chan, C. G., Yow, W. Q., & Oei, A. (2020). Active bilingualism in aging: Balanced bilingualism usage and less frequent language switching relate to better conflict monitoring and goal maintenance ability. *The Journals of Gerontology: Series B*, 75(9), e231–e241.
- 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(2), 338–348. <https://doi.org/10.1037/a0015225>
- Chen, S., Lin, Y., Zuo, S., Wang, Z., Liang, J., Jiang, Z., ... & Lin, L. (2022). Cognitive advantage of bilingualism over monolingualism in older adults: A meta-analysis. *Current Alzheimer Research*, 19(8), 555–567.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>

- Cox, S. R., Bak, T. H., Allerhand, M., Redmond, P., Starr, J. M., Deary, I. J., & MacPherson, S. E. (2016). Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia*, *91*, 299–306.
- Dass, R., Smirnova-Godoy, I., McColl, O., Grundy, J. G., Luk, G., & Anderson, J. A. (2024). A content overlap analysis of bilingualism questionnaires: Considering diversity. *Bilingualism: Language and Cognition*, 1–7. <https://doi.org/10.1017/S1366728923000767>
- De Lillo, M., & Ferguson, H. J. (2023). Perspective-taking and social inferences in adolescents, young adults, and older adults. *Journal of Experimental Psychology: General*, *152*(5), 1420–1438. <https://doi.org/10.1037/xge0001337>
- De Rezende, T. F., Bomfim, A. J. D. L., Chagas, N. M. D. S., Osório, F. D. L., & Chagas, M. H. N. (2018). Convergent validity of the Brazilian version of the theory of mind task battery for the assessment of social cognition in older adults. *Archives of Clinical Psychiatry (São Paulo)*, *45*, 75–76.
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. *International Review of Social Psychology*, *30*(1), 92–101.
- DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565–7574.
- DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2020). Duration and extent of bilingual experience modulate neurocognitive outcomes. *NeuroImage*, *204*, 116222.
- DeLuca, V., & Voits, T. (2022). Bilingual experience affects white matter integrity across the lifespan. *Neuropsychologia*, *169*, 108191.
- Dodell-Feder, D., Ressler, K., & Germine, L. (2020). Social cognition or social class and culture? On the interpretation of differences in social cognitive performance. *Psychological Medicine*, *50*(1), 133–145.
- Dotson, V. M., & Duarte, A. (2020). The importance of diversity in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 181–191.
- Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338.
- Fan, S.P., Liberman, Z., Keysar, B., & Kinzler, K.D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26* (7), 1090–1097.
- Feng, J., Cho, S., & Luk, G. (2023). Assessing theory of mind in bilinguals: A scoping review on tasks and study designs. *Bilingualism: Language and Cognition*, 1–15. <https://doi.org/10.1017/S1366728923000585>
- Fernandes, C., Barbosa, F., Martins, I. P., & Marques-Teixeira, J. (2021). Aging and social cognition: A comprehensive review of the literature. *Psychology & Neuroscience*, *14*(1), 1–15. <https://doi.org/10.1037/pne0000251>
- Ferreira Pereira, N. K., de Medeiros Cirne, G. N., de Oliveira Galvão, F. R., Costa, M. E., dos Santos Lima Junior, W., Azevedo Cacho, E. W., ... & de Oliveira Cacho, R. (2022). Reliability of the theory of mind task battery (ToM TB) to assess social cognition in post-stroke patients. *Topics in Stroke Rehabilitation*, *29*(7), 499–506.
- Fischer, A. L., O'Rourke, N., & Loken Thornton, W. (2017). Age differences in cognitive and affective theory of mind: Concurrent contributions of neurocognitive performance, sex, and pulse pressure. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(1), 71–81.
- Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2033–2039.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*, 531–534.
- Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, *15*, 512–519.
- Gallo, F., & Abutalebi, J. (2023). The unique role of bilingualism among cognitive reserve-enhancing factors. *Bilingualism: Language and Cognition*, *27*(2), 287–294.
- Gallo, F., DeLuca, V., Prystauka, Y., Voits, T., Rothman, J., & Abutalebi, J. (2022). Bilingualism and aging: Implications for (delaying) neurocognitive decline. *Frontiers in Human Neuroscience*, *16*, 22.
- Gallo, F., Novitskiy, N., Myachykov, A., Shtyrov, Y. (2021). Individual differences in bilingual experience modulate executive control network and performance: Behavioral and structural neuroimaging evidence. *Bilingualism: Language and Cognition*, *24*, 293–304.
- Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, *6*(1), 1–15. <https://doi.org/10.1017/S1366728903001007>
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multilingual naming test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *15*(3), 594–615.
- Grainger, S. A., Crawford, J. D., Riches, J. C., Kochan, N. A., Chander, R. J., Mather, K. A., Sachdev, P. S., & Henry, J. D. (2023). Aging is associated with multidirectional changes in social cognition: Findings from an adult life-span sample ranging from 18 to 101 years. *The Journals of Gerontology: Series B*, *78*(1), 62–72.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530.
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.
- Greenberg, D. M., Warrior, V., Abu-Akel, A., Allison, C., Gajos, K. Z., Reinecke, K., Rentfrow, P. J., Radecki, M. A., & Baron-Cohen, S. (2023). Sex and age differences in “theory of mind” across 57 countries using the English version of the “reading the mind in the eyes test”. *Proceedings of the National Academy of Sciences*, *120*(1), e2022385119.
- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, *117*, 123–134.
- Gullifer, J. W., & Titone, D. (2018). *Compute language entropy with {languageEntropy}*. Retrieved from <https://github.com/jasongullifer/languageEntropy>.
- Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283–294.
- Hamilton, L. J., Gourley, A. N., & Krendl, A. C. (2022). They cannot, they will not, or we are asking the wrong questions: Re-examining age-related decline in social cognition. *Frontiers in Psychology*, *13*, 894522.
- Happé, F. G., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology*, *34*(2), 358–362.
- Hartanto, A., & Yang, H. (2019). The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *149*, 609–633. <https://doi.org/10.1037/xge0000672>
- Henry, J. D., Grainger, S. A., & von Hippel, W. (2023). Determinants of social cognitive aging: Predicting resilience and risk. *Annual Review of Psychology*, *74*, 167–192.
- Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, *28*(3), 826–839.
- Hutchins, T. L., Prelock, P. A., & Chace, W. (2008). Test-retest reliability of a theory of mind task battery for children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *23*(4), 195–206.
- Jarvis, S. N., & Miller, J. K. (2017). Self-projection in younger and older adults: A study of episodic memory, prospection, and theory of mind. *Aging, Neuropsychology, and Cognition*, *24*(4), 387–407.
- Johansson Nolaker, E., Murray, K., Happé, F., & Charlton, R. A. (2018). Cognitive and affective associations with an ecologically valid test of theory of mind across the lifespan. *Neuropsychology*, *32*(6), 754.
- Kassambara, A. (2017). Practical guide to principal component methods in R (Vol. 2). Sthda.
- Kline, R.B. (2011). *Principles and practice of structural equation modeling* (5th ed.). New York: The Guilford Press.
- Kong, Q., Currie, N., Du, K., & Ruffman, T. (2022). General cognitive decline does not account for older adults' worse emotion recognition and theory of mind. *Scientific Reports*, *12*(1), 6808.
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, *12*, 48–54.
- Krendl, A. C., Mannerling, W., Jones, M. N., Hugenberg, K., & Kennedy, D. P. (2023). Determining whether older adults use similar strategies to young adults in theory of mind tasks. *The Journals of Gerontology: Series B*, *78*(6), 969–976.

- Lee, C. K., Collinson, S. L., Feng, L., & Ng, T. P. (2012). Preliminary normative neuropsychological data for an elderly Chinese population. *The Clinical Neuropsychologist*, *26*(2), 321–334.
- Li, X., Ng, K. K., Wong, J. J. Y., Lee, J. W., Zhou, J. H., & Yow, W. Q. (2021). Bilingual language entropy influences executive functions through functional connectivity and signal variability. *Brain and Language*, *222*, 105026.
- Li, X., Ng, K. K., Wong, J. J. Y., Zhou, J. H., & Yow, W. Q. (2024). Brain gray matter morphometry relates to onset age of bilingualism and theory of mind in young and older adults. *Scientific Reports*, *14*(1), 3193.
- Lorge, I., & Katsos, N. (2019). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approaches to Bilingualism*, *9*(3), 376–397. <https://doi.org/10.1075/lab.16054>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621.
- Luk, G., Bialystok, E., Craik, F. I., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *Journal of Neuroscience*, *31*(46), 16808–16813.
- Luk, G., Mesite, L., & Guerrero, S. L. (2020). Onset age of second language acquisition and fractional anisotropy variation in multilingual young adults. *Journal of Neurolinguistics*, *56*, 100937. <https://doi.org/10.1016/j.jneroling.2020>.
- McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.
- Murphy, J., Millgate, E., Geary, H., Catmur, C., & Bird, G. (2019). No effect of age on emotion recognition after accounting for cognitive factors and depression. *Quarterly Journal of Experimental Psychology*, *72*(11), 2690–2704.
- Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699.
- Navarro, E. (2022). What is theory of mind? A psychometric study of theory of mind and intelligence. *Cognitive Psychology*, *136*, 101495.
- Navarro, E., & Conway, A. R. (2021). Adult bilinguals outperform monolinguals in theory of mind. *Quarterly Journal of Experimental Psychology*, *74*(11), 1841–1851.
- Navarro, E., DeLuca, V., & Rossi, E. (2022). It takes a village: Using network science to identify the effect of individual differences in bilingual experience for theory of mind. *Brain Sciences*, *12*(4), 487. <https://doi.org/10.3390/brainsci12040487>
- Ng, T. P., Feng, L., Lim, W. S., Chong, M. S., Lee, T. S., Yap, K. B., Tsoi, T., Liew, T. M., Gao, Q., Collinson, S., Kandiah, N., & Yap, P. (2015). Montreal cognitive assessment for screening mild cognitive impairment: Variations in test performance and scores by education in Singapore. *Dementia and Geriatric Cognitive Disorders*, *39*(3–4), 176–185.
- Nguyen, T. K., & Astington, J. W. (2014). Reassessing the bilingual advantage in theory of mind and its cognitive underpinnings. *Bilingualism: Language and Cognition*, *17*(2), 396–409.
- Ooi, S. H., Goh, W. D., Sorace, A., & Bak, T. H. (2018). From bilingualism to bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition*, *21*(4), 867–879.
- Osterhaus, C., & Bosacki, S. L. (2022). Looking for the lighthouse: A systematic review of advanced theory-of-mind tests beyond preschool. *Developmental Review*, *64*, 101021. <https://doi.org/10.1016/j.dr.2022.101021>
- Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child Development*, *87*(6), 1971–1991.
- Osterhaus, C., & Koerber, S. (2021). The development of advanced theory of mind in middle childhood: A longitudinal study from age 5 to 10 years. *Child Development*, *92*(5), 1872–1888. <https://doi.org/10.1111/cdev.13627>
- Otsuka, Y., Nakai, R., Shizawa, M., Itakura, S., Sato, A., & Abe, N. (2024). Brain structure variation and individual differences in theory of mind among older adults. *Aging Brain*, *5*, 100115. <https://doi.org/10.1016/j.nbas.2024.100115>
- Papageorgiou, A., Bright, P., Periche Tomas, E., & Filippi, R. (2019). Evidence against a cognitive advantage in the older bilingual population. *Quarterly Journal of Experimental Psychology*, *72*(6), 1354–1363.
- Perani, D., Farsad, M., Ballarini, T., Lubian, F., Malpetti, M., Fracchetti, A., Magnani, G., March, A., & Abutaleb, J. (2017). The impact of bilingualism on brain reserve and metabolic connectivity in Alzheimer's dementia. *Proceedings of the National Academy of Sciences*, *114*(7), 1690–1695. <https://doi.org/10.1073/pnas.1610909114>
- Phillips, L. H., Bull, R., Allen, R., Insch, P., Burr, K., & Ogg, W. (2011). Lifespan aging and belief reasoning: Influences of executive function and social cue decoding. *Cognition*, *120*(2), 236–247.
- Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*, 127–152. <https://doi.org/10.1080/17405620344000022>
- Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, *15*(2), 384–396.
- Rahman, F. (2021). *The neuro-cognitive profile of theory of mind in healthy ageing* [Doctoral dissertation, Aston University].
- Rahman, F., Kessler, K., Apperly, I. A., Hansen, P. C., Javed, S., Holland, C. A., & Hartwright, C. E. (2021). Sources of cognitive conflict and their relevance to theory-of-mind proficiency in healthy aging: A preregistered study. *Psychological Science*, *32*(12), 1918–1936.
- Raimo, S., Cropano, M., Roldán-Tapia, M. D., Ammendola, L., Malangone, D., & Santangelo, G. (2022). Cognitive and affective theory of mind across adulthood. *Brain Sciences*, *12*(7), 899.
- Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review*, *24*(3), 355–370.
- Roheger, M., Brenning, J., Riemann, S., Martin, A. K., Flöel, A., & Meinzer, M. (2022). Progression of socio-cognitive impairment from healthy aging to Alzheimer's dementia: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *140*, 104796. <https://doi.org/10.1016/j.neubiorev.2022.104796>
- Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 211–217. <https://doi.org/10.1037/a0025162>
- Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language*, *74*, 46–76.
- Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study. *Cognitive Science*, *42*(7), 2313–2341.
- Sandoz, M., Démonet, J. F., & Fossard, M. (2014). Theory of mind and cognitive processes in aging and Alzheimer type dementia: A systematic review. *Aging & Mental Health*, *18*(7), 815–827.
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65–72.
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*, 442717.
- Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook* (Vol. 17). Western Psychological Services.
- Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Frontiers in Communication*, *3*, 36. doi: 10.3389/fcomm.2018.00036
- Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology*, *24*(5), 591–601.
- Schweizer, T. A., Ware, J., Fischer, C. E., Craik, F. I., & Bialystok, E. (2012). Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. *Cortex*, *48*(8), 991–996.
- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*(3), 448–460.
- Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028.
- Stern, Y., Barnes, C. A., Grady, C., Jones, R. N., & Raz, N. (2019). Brain reserve, cognitive reserve, compensation, and maintenance: Operationalization, validity, and mechanisms of cognitive resilience. *Neurobiology of Aging*, *83*, 124–129.

- Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge/Taylor & Francis Group.
- Stevens, W. D., Khan, N., Anderson, J. A., Grady, C. L., & Bialystok, E. (2023). A neural mechanism of cognitive reserve: The case of bilingualism. *NeuroImage*, **281**, 120365. <https://doi.org/10.1016/j.neuroimage.2023.120365>
- Sulpizio, S., Del Maschio, N., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2020). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *NeuroImage*, **205**, 116306.
- Tiv, M., O'Regan, E., & Titone, D. (2021). In a bilingual state of mind: Investigating the continuous relationship between bilingual language experience and mentalizing. *Bilingualism: Language and Cognition*, **24**(5), 918–931.
- Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, **22**(3), 516–536.
- Wang, Z., & Su, Y. (2013). Age-related differences in the performance of theory of mind in older adults: A dissociation of cognitive and affective components. *Psychology and Aging*, **28**(1), 284–291. <https://doi.org/10.1037/a0030876>
- Ware, A. T., Kirkovski, M., & Lum, J. A. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, **11**, 1458.
- Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, **191**, 103997.
- Wechsler, D. (2000). *Manuel de la WAIS-III*. Paris: Centre de Psychologie Appliquée.
- 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.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, **75**(2), 523–541.
- Wilson, R. S., Boyle, P. A., Yu, L., Barnes, L. L., Schneider, J. A., & Bennett, D. A. (2013). Life-span cognitive activity, neuropathologic burden, and cognitive aging. *Neurology*, **81**(4), 314–321.
- Yeung, E. K. L., Apperly, I. A., & Devine, R. T. (2024). Measures of individual differences in adult theory of mind: A systematic review. *Neuroscience & Biobehavioral Reviews*, 105481.
- Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in Psychology*, **6**, 164.
- Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental Science*, **20**(1), e12482.
- Yu, C. L., Kovelman, I., & Wellman, H. M. (2021). How bilingualism informs theory of mind development. *Child Development Perspectives*, **15**(3), 154–159.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, **27**(8), 1–25.
- Zuur A.F., Ieno E.N., Elphick C.S. (2010). A protocol for data exploration to avoid common statistical problems: Data exploration. *Methods in Ecology and Evolution*, **1**, 3–14.