

Sex and age interact in reading the mind in the eyes

Matthias L. Schroeter^{a,b,c,*}, Jana Kynast^{a,c}, Haiko Schlögl^d, Simon Baron-Cohen^e,
Arno Villringer^{a,b,c}

^a Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurology, Leipzig, Germany

^b University Hospital Leipzig, Clinic for Cognitive Neurology, Leipzig, Germany

^c University of Leipzig, LIFE - Leipzig Research Center for Civilization Diseases, Leipzig, Germany

^d Division of Endocrinology, Department of Endocrinology, Nephrology, Rheumatology, University Hospital Leipzig, Leipzig, Germany

^e University of Cambridge, Department of Psychiatry, Autism Research Centre, Cambridge, UK

ARTICLE INFO

Keywords:

Aging
Mindreading
Reading the mind in the eyes test
Sex
Social cognition

ABSTRACT

Social cognition includes understanding the mental states (thoughts, feelings, intentions, desires, and beliefs) of others – so-called ‘theory of mind’ or ‘mindreading’. Recent studies have shown an impact of age and sex. Here, we applied the ‘Reading the Mind in the Eyes’ Test (RMET) that measures the ability to identify mental states from the eye region of the face. RMET accuracy was measured and analyzed in a large population-based sample ($N = 1603$) across the whole adult age-range from 19 to 79 years with effect size analyses (Hedges’ g). Overall test performance was lower in older than younger women and men, whereas differences between women and men were almost negligible across the whole cohort. In a further analysis focusing on age-specific sex differences, RMET accuracy was higher for women below 45 years compared to men. This sex effect nearly vanished in older people above 45 years of age. Results were verified in a sub-cohort after excluding participants with neurological and psychiatric conditions, and with another cut-off, i.e. 50 years of age. In conclusion, results suggest that mindreading declines with age. Overall sex effects were small and results suggest that age-related hormonal and social factors may impact mental state perception. Future mega-analyses and longitudinal studies including hormonal and social measures are needed to validate the interaction between RMET performance, aging and sex.

1. Introduction

Theory of Mind (ToM) also known as ‘mentalizing’ or ‘mindreading’ is a crucial socio-cognitive skill [1,2]. It enables the attribution of mental states to self and others and, herewith, adequate social communication and interaction. A well-known and frequently used test to investigate social cognition is the ‘Reading the Mind in the Eyes’ Test (RMET), where people attribute mental states to gaze [3,4]. Here, participants select the term that most appropriately describes the mental state of the person in photographs of the eye-region.

Recently, we investigated RMET accuracy in the context of age and cognition across the whole adult age-range (19–79 years) in a large population-based sample ($N = 1603$) [2]. In line with an aging-focused meta-analysis of social cognition [5], lower RMET accuracy was significantly associated with advanced age, indicating problems inferring mental state from gaze at older age.

Sex (i.e., male or female biological sex) is a key variable impacting social cognitive processes directly and/or indirectly by interaction with

related physiological and environmental variables (e.g., brain, genes, hormones, social experience, and culture; [6,7]). Recent findings regarding RMET performance are contradictory. While some studies do not find differences between men and women regarding RMET accuracy rates [8–10], others hint at a small but robust advantage for women over men [4,9,11–16]. Critically, past studies investigating these sex effects have been performed mainly in highly selective study samples (i.e. highly educated, young adults/undergraduates). Based on meta-analytic evidence [4] and our own studies [2,17], we expected a small on-average performance advantage in women relative to men. In line with this, RMET performance has been reported to be sensitive to hormone effects, including testosterone and estrogen [18–21].

Having a large lifespan-cohort of men and women available for testing allowed us to investigate whether any sex differences in RMET performance are still detectable after the typical age of menopause. The time before menopause, i.e. perimenopause, is a phase of heightened fluctuations in endogenous sex steroid levels, followed by follicular depletion and low concentrations of sex steroids estrogen, progesterone

* Corresponding author. Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A, 04103, Leipzig, Germany.

E-mail address: schroet@cbs.mpg.de (M.L. Schroeter).

<https://doi.org/10.1016/j.cpnec.2022.100162>

Received 17 August 2022; Received in revised form 4 November 2022; Accepted 8 November 2022

Available online 9 November 2022

2666-4976/Crown Copyright © 2022 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and testosterone during the postmenopausal years [22]. Since, on average, women enter the menopause approximately by 45 years of age [23], this age was set as a proxy for menopausal status to detect differences between men and women as a function of hormonal variations pre- versus post-menopause. We hypothesized that the sex-effect on RMET outcomes decreased after menopause, suggesting that higher estrogen and/or progesterone concentrations have a positive effect on test performance and mediate the sex differences observed in younger persons. However, since we do not directly measure hormonal status, our main analysis centers on testing for an age by sex interaction.

2. Materials and methods

This study was part of the population-based study of the Leipzig Research Centre for Civilization Diseases [24]. Of randomly sampled people, where data had been provided by the resident's registration office of the city of Leipzig, 10,000 participants were tested in total from August 2011 to November 2014. Most participants were of European ancestry. A subsample was characterized by in-depth cognitive phenotyping including the RMET (final cohort 1603 persons; age range 19–79 years). Detailed information can be found in Ref. [2].

ToM was assessed with the German version of the revised RMET [3, 25]. It includes 36 pictures showing the eye region. Every picture is presented with four words of which the best fitting mental state has to be selected. The test was self-paced and took 10–15 min to complete. Accuracy rates (percentage of correct responses) were analyzed. All individuals scored above chance level (>25%), ensuring sufficient understanding of the test.

Information about sex was obtained directly from the resident's registration office of the city of Leipzig for each participant, at the time of data collection the possibilities were male and female. Moreover, participants filled in a questionnaire with the question: "What is your gender?" ("Welches Geschlecht haben Sie?"). Options were male or female. In 93.4% of the study population, information on both factors was available and matched in all cases.

Women were classified as likely premenopausal if they were younger than 45 years or likely peri- or postmenopausal if they were 45 years and older, following [23]. As mentioned above, age was used as a proxy for menopausal status in women, as objective information about hormonal concentrations or anamnestic information about menstruation was not available. The study was approved by the ethics committee of the University of Leipzig and was in accordance with the latest version of the Declaration of Helsinki. Each participant provided written informed consent.

To analyze differences between women and men in RMET performance, Hedges' g was calculated [26]. As a standardized measure Hedges' g was chosen as effect size measure because it corrects for differences in sample size [27]. According to established conventions [28], effects can be classified as small $g = |0.2|$, medium $g = |0.5|$, or large $g = |0.8|$. Magnitude of the difference in RMET accuracy between women and men was calculated in the whole sample and within each 5-year age group (19–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–80 years), performance of men was referenced to performance of women. Thus, negative scores indicate lower accuracy for men relative to women. To investigate the effect of possible menopausal state, effect sizes for the sex effect were additionally calculated for individuals younger and equal or older than 45 years of age. In all cases, respective 95% confidence intervals of the effect sizes were calculated. To test differences between men and women in young (<45 years) compared to older adults (≥ 45 years), effect sizes were compared with an independent sample t -test. Based on the values published in a recent meta-analysis ($g = 0.177$, $\beta = 0.2$, $\alpha = 0.01$; [4]), a power analysis was conducted to estimate if our sample was large enough to detect differences between men and women in the RMET.

To calculate the size of the age effect separately for men and women, the youngest age group (19–24 years) served as reference for any other

age group. In both men and women, we expected the performance gap relative to the youngest age group to increase with older age. Additionally, the age groups below and above 45 years were contrasted for both sexes. As the younger age group served as reference group, negative effect sizes indicate lower performance in the age group equal or older than 45 years.

Analyses were repeated in a smaller subsample after excluding participants with neurological and psychiatric conditions (exclusion criteria see Ref. [2]), finally consisting of 978 participants (576 men), to ensure validity of results.

3. Results

There was no age difference between the 698 women (60.4 ± 14.0) and 905 men (60.3 ± 15.0 years) for the whole cohort ($t[1; 1544] = -0.118$, $p > 0.05$; mean \pm standard deviation, t [degrees of freedom; n]). Women and men performed with slight advantage of women, but statistically missing the threshold regarding accuracy in the RMET ($63.8 \pm 10.5\%$ vs. $62.8 \pm 10.5\%$; $t(1; 1601) = -1.956$, $p = 0.051$).

Fig. 1A shows the effect sizes for RMET performance of men relative to women for the whole sample and the single age groups. According to established conventions [28], the overall effect of sex ($g = -0.1$) can be interpreted as very small in accordance with the aforementioned group comparison. Within the single age groups, effect sizes can be classified as small but few numbers also fall below the predefined range of small effects ($g < |0.2|$, [28]; range $g = -0.25$ to $g = 0.127$). Only for two of the younger age groups, i.e., 40–44 years ($g = -0.53$) and 30–34 years ($g = -0.44$), medium effect sizes were found. Fig. 1 shows that the sex difference is largest for the groups below the age of 45 years (mean $g = -0.29$). Here, women outperformed men by roughly half a standard deviation, reflecting an accuracy advantage of 5%. Older age groups presented smaller effect sizes for sex differences (mean $g = -0.08$) reflecting an accuracy advantage of 2% only. The effect of sex was significantly larger for individuals younger than 45 years compared to older adults ($t(10) = -2.189$, $p < 0.05$ one-tailed).

Fig. 1B shows Hedges' g for men and women in distinct age groups. When referencing RMET performance to the youngest age group (19–24 years), Hedges' g scores linearly declined at older age for men and women alike. Interestingly, a difference in age-related effect sizes between men and women can be observed. The age effect is most dissimilar for men and women below the age of 45 years. Effect sizes are larger and slightly more negative for men compared to women of the same age. For older age groups, effect sizes for males and females are in a similar range. Starting at the age of 45 years, effect sizes for both men and women can be classified as large (range -1.3 to -0.55). If effect size was calculated for the decline in RMET performance above age of 45 years vs. below this age as a threshold, the effect was large in women ($g = -0.75$), but only medium-sized in men ($g = -0.55$) (see Fig. 1, two bottom bars), which corresponds with pronounced sex differences with an advantage for women below 45 years of age.

Effect size analyses performed on the selected subsample without neurological diseases or psychiatric disorders ($N = 978$) revealed generally comparable results to analyses performed on the entire sample. Respective results are presented in Fig. 2. Accordingly, results of the first analysis were confirmed and were not related to possibly confounding factors such as neurological diseases or psychiatric disorders. Note that the effect size analysis also proved the significantly larger sex effect in individuals younger than 45 years compared to older adults.

4. Discussion

We investigated mindreading from the eyes in a population-based lifespan sample. The study included more than 1600 randomly selected individuals. The primary aim was assessing differences between women and men for several age groups and, in particular, the potential impact of likely menopausal state.

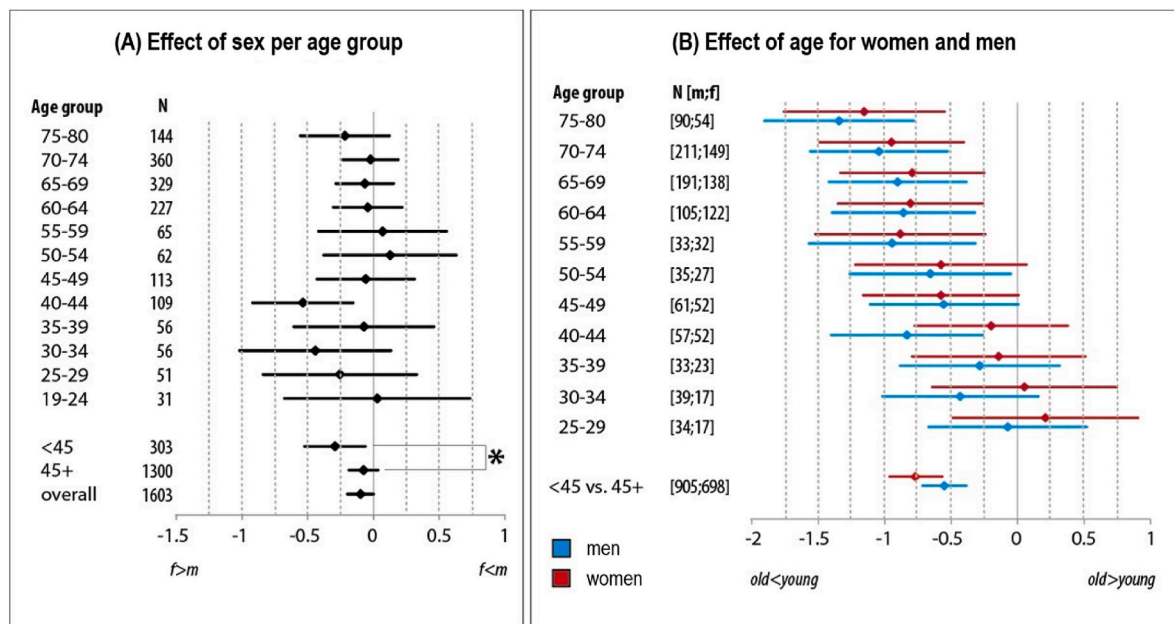


Fig. 1. Impact of sex and age on performance in the Reading the Mind in the Eyes Test – Effect size analysis across the whole sample ($N=1603$). **(A)** Hedges' g is displayed as performance of men relative to performance of women for all depicted groups. Negative values indicate worse performance for men compared to women. *Difference in the mean effect sizes of the sex effect between age groups below and above (\geq) 45 years was tested with independent sample t -test ($t(10) = -2.189$; $p < 0.05$ one tailed). **(B)** Hedges' g for every age group relative to the age group 19–24 years is shown for men and women. Negative values indicate worse performance of an age group compared to the youngest age group. The bottom line depicts the comparison between the age groups above (\geq) and below 45 years for men and women. Note: N number of individuals, m male and f female subjects.

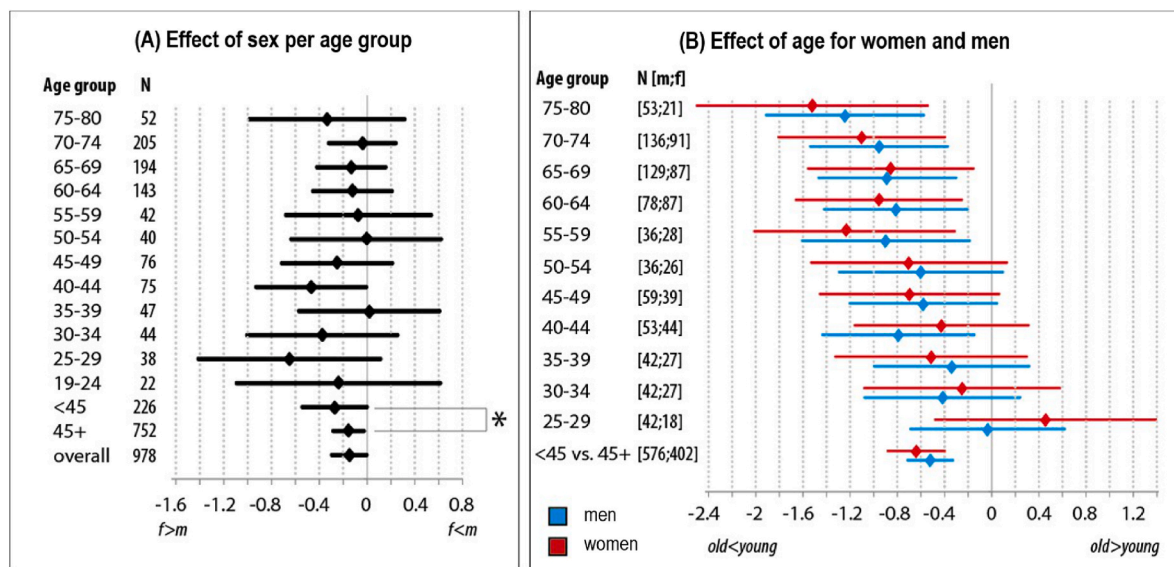


Fig. 2. Impact of sex and age on performance in the Reading the Mind in the Eyes Test – Effect size analysis in the subsample without neurological diseases or psychiatric disorders ($N=978$). **(A)** Hedges' g is displayed as performance of men relative to performance of women for all depicted groups. Negative values indicate worse performance for men compared to women. *Difference in the mean effect sizes of the sex effect between age groups below and above (\geq) 45 years was tested with independent sample t -test ($t(10) = -1.928$; $p < 0.05$ one-tailed). **(B)** Hedges' g for every age group relative to the age group 19–24 years is shown for men and women. Negative values indicate worse performance of an age group compared to the youngest age group. The bottom line depicts the comparison between the age groups above (\geq) and below 45 years for men and women. Note: N number of individuals, m male and f female subjects.

4.1. Aging impairs reading the mind from the eyes

First of all, we detected an age effect in the RMET, i.e., better performance in younger compared with older women and men, in line with our previous analyses in the same dataset [2] and an independent meta-analysis [5]. Effect sizes decreased in both men and women at older age when comparing RMET performance to the youngest age

group. Taking the age of 45 years as a cut-off [23], the age effect in males can be interpreted as medium-sized, while it is large in women. Given a pooled standard deviation of 1, older women detected emotional content in the RMET 7.5% (2.7 items) less accurately compared to younger women. In men, elderly scored 5.5% (~1.98 items) worse compared to younger men. In sum, our results indicate an age-related decline in RMET performance, slightly more pronounced in women than men. A

similar trend was found in a recent, large-scale study on emotion recognition using dynamic stimuli [29].

4.2. Women slightly outperform men in reading the mind from the eyes

Regarding sex, effect sizes suggested a small overall advantage for women over men. Based on the numbers provided in a recent meta-analysis [4], we conducted a power analysis, resulting in a minimal number of 320 participants per group to obtain appropriate statistical power for the detection of significant differences between men and women. Our sample exceeds this number crucially. The sex effect we found, i.e., $g = -0.1$, is slightly smaller compared to the meta-analysis by Ref. [4] with $g = 0.177$ (note that positive scores reflect female superiority here, since males served as reference group), which might be related to different age distributions. Effect sizes obtained in our study are comparable with another large-scale study [16]. This study used the RMET as well and included approximately 90,000 research volunteers of European ancestry. Here, the female advantage on the RMET reached comparable values (Cohen's $d = 0.21$). In sum, our data indicate only slight general differences between men and women in mindreading from the eyes when analyzing the whole sample.

4.3. Sex and age interact in reading the mind from the eyes

Age-specific patterns of the sex effect emerged from a more detailed investigation. Results support a female performance advantage over males in age groups younger than 45 years, in contrast to age groups equal to or older than 45 years where this sex difference disappears. Translated to accuracy rates, young women performed 5% better compared to men, whereas older women outperformed men by 2% only. This finding coincides with the stronger age-related decline in RMET performance in women compared to men.

Results are in line with our hypothesis and might indicate a potential modulating effect of sex steroid hormones as perimenopausal transition occurs approximately at an age of 45 years in women with fluctuating ovarian hormone production and subsequent cessation of ovarian function [23]. Perimenopause has been discussed as 'a critical period in the neuroadaptive landscape of ageing in the female brain' [22]. It is a midlife transition state that leads to reproductive senescence in women and has been discussed to be accompanied by emotional and cognitive dysfunction. Furthermore, it is associated with a decline in glucose metabolism in the brain, in particular in regions rich in estrogen receptors, and an increased risk for neurodegenerative diseases, particularly Alzheimer's disease [22].

One might argue that our cut-off value of 45 years is too restrictive as between onset of perimenopause and menopause estrogen production still occurs, and because menopause has individually a wide age range from 40 to 58 years [22]. Although a large meta-analysis spanning 35 countries identified an overall mean age of menopause at 48.8 years, it shows a relevant variation by geographical region. Whereas values for the United States of America and Asia were close to this mean value, age at menopause was generally lower in Africa, Latin America and Middle-Eastern countries (47.2–48.4 years), but higher in Europe and Australia (50.5–51.2 years) [30]. Accordingly, we reanalyzed the interaction effect between sex and aging with a cut-off value of 50 years. Again, the sex differences were larger for the groups below the age of 50 years (mean $g = -0.22$) with better performance in women than men, but almost vanished in older age groups (mean $g = -0.08$). The effect of sex was significantly larger for individuals younger than 50 years compared to older adults ($p = 0.037$ one-tailed for the whole group; $p = 0.067$ one-tailed for the selected group).

On the other hand, this hypothesis – focusing on perimenopause or menopause – might be an oversimplification, especially with lacking hormonal data or self-reported menopausal status. Interaction effects between sex and aging in RMET performance might also be related to aging-related decline of testosterone concentrations in men in our study

that is associated with cognitive impairment [31]. Testosterone has been shown to impair social cognition in the RMET [18–21]. In men, testosterone concentrations decrease gradually approximately with 1% testosterone loss per year after 30 years of age [32]. Prior to menopause, women have high estrogen concentrations in blood. During perimenopause, the sex hormone profile changes due to the loss of ovarian function to very low estradiol concentrations, with maintained ovarian and adrenal androgen production. In men, estrogens are formed by aromatization of androgens, a process that occurs lifelong. Thus, aging men have higher estrogen blood concentrations than aging women. Consequently, interaction between aging and sex in RMET performance, as observed in our study, might be related to hormonal reconfiguration in both sexes with aging.

Beside hormonal changes, social factors may also influence the observed interaction effect as environmental factors are likely to modulate the relation between sex, age and mentalizing abilities. In particular, social standards that are established in the behavioral repertoire of members of a society are passed on to the next generation by key mechanisms described in social learning theory [33] and social cognitive theory [34]. Thus, social values leave an imprint on a person's personality and cognitive development. As Germany has been divided into the socialist part in the east and capitalistic part in the west between 1949 and 1990, the majority of older participants in our sample was socialized and gained social cognitive abilities in a socialist environment, while individual development in younger participants was predominantly influenced by social values originating in a capitalistic culture. Thus, it cannot be completely ruled out that the difference in the sex effect on mindreading between younger and older adults might be modulated by differences in social environment during development. More recent findings further support the impact of socio-structural factors on social cognitive development [35,36].

In conclusion, we report subtle but significant sex-differences in age-related RMET performance in effect size analyses contrasting age-groups. These data might indirectly suggest a modulating role of sex hormone concentration and social factors in cognitive empathy/mindreading. Our results might encourage thorough future studies addressing these factors more carefully, in particular with inclusion of hormonal and social parameters.

4.4. Mind reading ability increases in childhood to early adulthood, but decreases in higher age

Although our recent analysis [2] based on the same dataset revealed comparable findings for aging, it generally did not show significant effects of sex or a significant interaction between sex and aging. One might argue that these results contradict our current findings. However, [2] applied linear regression analyses to investigate influences of age, sex and verbal intelligence on RMET performance. A further analysis included additionally potentially modulating factors, i.e. cognitive performance in the cognitive domains attention, executive function, memory, and verbal fluency. Linear regression analyses performed on the selected subsample without neurological or psychiatric conditions revealed generally comparable results to analyses performed on the entire sample, i.e., it did not show significant effects of sex or for the interaction between sex and age. In sum, the linear regression analyses did neither reveal effects of sex nor an interaction between sex and aging, although conducted in the same dataset. However, linear regression analyses are sensitive to linear effects across aging only. A difference between women and men in RMET performance changing with a certain age (here 45 or 50 years) cannot be detected with such linear approaches. In contrast, effect size analyses are sensitive to detect such alterations. One might conclude that results do not contradict, rather complement each other. In detail, the effect size analysis as presented in the current study confirms the age effect already observed in the former study by Ref. [2]; and adds the impact of sex on age-related changes in the RMET as a new finding. Both statistical approaches are

justified by answering different research questions.

Other comprehensive studies also investigated RMET performance in relation to age and sex. [37] investigated 902 adult participants between 66 and 105 years with a 10-item version of the RMET. Besides calculating normative scores, the authors showed better performance in the RMET with younger age, higher education, white ethnicity, higher cognitive screening scores, literacy, social norms scores, and higher scores in all five domains in cognitive composites. RMET scores were lower in people with higher depression and anxiety symptoms.

Another comprehensive study by Ref. [38] evaluated the impact of demographic and sociocultural factors (age, gender, education, and ethnicity) on RMET performance and other social and non-social cognitive tasks. They included 40,248 international, native-/primarily English-speaking participants between the ages of 10 and 70 years. Participants completed the RMET, a shortened version of the RMET, a multiracial emotion identification task, an emotion discrimination task, and a non-social/non-verbal processing speed task (digit symbol matching). Contrary to other tasks, performance on the RMET improved across the lifespan. The authors revealed that education and ethnicity explained more variance in RMET performance than the other tasks, where more highly educated, non-Hispanic, and White/Caucasian individuals performed best. They concluded that the RMET may be unduly influenced by social class and culture.

At first glance, results by Refs. [37,38] might question or even be inconsistent with our findings for aging. Whereas our study investigated an age-range from 19 to 79 years, [37] focused on an older cohort between 65 and 105, and [38] on a younger cohort from 10 to 70 years. One might hypothesize that performance in the RMET increases in earlier adult life, in particular from pre-adolescence to early adulthood to an age of 18–20 years (see Fig. 1a by Ref. [38]), but decreases in higher age beyond 60–70 years. Note that [38] data indicate a decline in RMET performance beyond an age of 65 years (see their Fig. 1a), where their linear analysis was most probably not sensitive to detect this effect.

Besides an effort to synthesize findings of these studies – best done in a pooled mega-analysis – differences might also be explained by differences in design and stimulus material. Whereas our cross-sectional study was population-based and done in person such as [37], [38] used an online internet approach. The latter design will have led most probably to a selection bias, in particular in the elderly not experienced with online studies. One might assume that well-educated elderly with high proficiency in internet usage are overrepresented in those data, leading to artificially high RMET scores in age. Secondly, different RMET version were applied, i.e., a 36-item RMET version by Ref. [2]; a 16- and a 36-item RMET version by Ref. [38]; but a 10-item RMET version by Ref. [37].

Finally, regarding the impact of sex, [38] demonstrated that women outperformed men with a low effect size of $d = 0.26$ in the 16-item RMET, well known from other studies (see above), but did not perform an analysis on the interaction between sex and aging as done in our study. The same holds true for [37]; who did not detect a significant impact of sex in the whole group, but in one subsample, i.e., their youngest group between 65 and 74 years of age with women outperforming men. As a desideratum for the future, a mega-analysis of these three studies might be conducted to validate our findings on the interaction between aging and sex.

4.5. Limitations and perspectives

Here we discuss possible limitations of our study. In our former study applying linear regression analyses [2], we could show that performance in the RMET is related to age, and beyond that, also to verbal intelligence as measured with the vocabulary test ('Wortschatztest', WST). Here, older age and lower verbal intelligence were significantly related to lower RMET accuracy. Of note, this study used the same dataset like our current study. Remarkably, individual performance in cognition, i.e. the cognitive domains attention, executive function, verbal fluency and

memory, did not have a consistent impact on RMET performance, minimizing the opportunity of a bias in our current study. Although we would have wished including the only relevant factor in our current study, i.e. verbal intelligence, the effect size analysis did not allow including this parameter as a control factor. Moreover, we could not analyze the impact of ethnicity, as our cohort included almost exclusively participants of European ancestry.

We applied a conventional binary concept of sex, i.e., identification with female or male sex. Our approach enables comparability with previous studies but further biological and environmental factors might impact individual self-concepts of gender identity (e.g., social roles, experiences, etc.; see Ref. [7]), which were not investigated in our study and shall be addressed in future studies including broader definitions beyond the conventional concept of sex/gender to fully address the diversity in gender identity and its impact on cognitive processes [2]. A further limitation is the use of age as a proxy for menopausal status. Accordingly, our study can only provide indirect evidence for the modulating role of perimenopausal transition on RMET performance. It remains an open issue for future research if and to what extent natural differences in sex hormones modulate the ability to identify mental states from gaze. Further longitudinal studies including detailed hormonal measures and/or a comprehensive characterization of reproductive ageing according to the 2011 Stages of Reproductive Aging Workshop (STRAW+10) staging system [39] will be needed to validate this finding. Furthermore, with such a thorough study design, hormonal changes in men with aging might be taken into account additionally. Finally, the number of participants was not balanced across age groups, an issue related to the conceptual development of the LIFE study focusing mainly on the elderly. However, numbers in younger age groups are still large compared to other studies, ensuring reliability of our findings. Additionally, we applied Hedges' g as an effect size measure as it adjusts for differences in sample size [27] and adds valuable information about the relevance of group differences.

In summary, our study investigated the effects of sex on mental state recognition from gaze in the RMET in a large population-based sample including 1603 adults aged 19–79 years, i.e., across the whole lifespan. Results reveal declining performance in the RMET with aging for both women and men. Although differences between women and men were small across the whole study cohort, results suggest a specific age pattern in men's and women's mindreading from the eyes, representing a larger performance advantage for women over men in younger age that vanishes with aging, presumably related to age-related hormonal changes and potentially modulated by social factors. Results of our cross-sectional study should be confirmed in future mega-analyses and longitudinal studies also involving other languages and ethnicities, including a comprehensive conceptualization of sex/gender and verifying relation to hormonal state and social parameters by in-depth assessment.

Author contributions

JK planned and conducted statistical analyses and created figures, and wrote the first draft of the manuscript together with MLS. AV and HS interpreted results and reviewed the final draft of the manuscript. SBC contributed substantially to the interpretation of the results. MLS designed the study, supervised data acquisition and analyses, substantially contributed to the interpretation of the results and wrote the manuscript together with JK. MLS also had the lead in revising the paper. All authors approved the final version of the manuscript.

Declaration of competing interest

The authors state that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This study has been supported by LIFE – Leipzig Research Center for Civilization Diseases, University Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by means of the Free State of Saxony within the framework of the excellence initiative (project numbers 713–241202, 713–241202, 14505/2470, 14575/2470). Part of this work had been supported by the German Research Foundation (DFG; CRC 1052 “Obesity mechanisms”, Project A1, to AV, and SCHR 774/5–1 to MLS). MLS additionally received funding from the eHealthSax Initiative of the Sächsische Aufbaubank (SAB; project TelDem). Accordingly, this study is co-financed with tax revenue based on the budget approved by the Saxon state parliament. We thank all participants and the team at the LIFE study centre, who made this study possible. JK was supported by the Max Planck International Research Network on Aging. SBC received funding from the Wellcome Trust (214322/Z/18/Z), the Innovative Medicines Initiative 2 Joint Undertaking (777394, project AIMS-2-TRIALS) with support from the European Union’s Horizon 2020 research and innovation programme, from the Autism Centre of Excellence, the Templeton World Charitable Fund, the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre and from the NIHR Applied Research Collaboration East of England.

References

- [1] P.J. Eslinger, S. Anders, T. Ballarini, S. Boutros, S. Krach, A.V. Mayer, J. Moll, T. L. Newton, M.L. Schroeter, R. de Oliveira-Souza, J. Raber, G.B. Sullivan, J.E. Swain, L. Lowe, R. Zahn, The neuroscience of social feelings: mechanisms of adaptive social functioning, *Neurosci. Biobehav. Rev.* 128 (2021) 592–620, <https://doi.org/10.1016/j.neubiorev.2021.05.028>.
- [2] J. Kynast, E.M. Quinke, M. Polyakova, T. Luck, S.G. Riedel-Heller, S. Baron-Cohen, A. Hinz, A.V. Witte, J. Sacher, A. Villringer, M.L. Schroeter, Mindreading from the eyes declines with aging - evidence from 1,603 subjects, *Front. Aging Neurosci.* 12 (2020), 550416, <https://doi.org/10.3389/fnagi.2020.550416>.
- [3] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, The “Reading the Mind in the Eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism, *JCPP (J. Child Psychol. Psychiatry)* 42 (2) (2001) 241–251, <https://doi.org/10.1017/s0021963001006643>.
- [4] R.A. Kirkland, E. Peterson, C.A. Baker, S. Miller, S. Pulos, Meta-analysis reveals adult female superiority in “reading the mind in the eyes” test, *N. Am. J. Psychol.* 15 (1) (2013) 121–146.
- [5] J.D. Henry, L.H. Phillips, T. Ruffman, P.E. Bailey, A meta-analytic review of age differences in theory of mind, *Psychol. Aging* 28 (3) (2013) 826–839, <https://doi.org/10.1037/a0030677>.
- [6] V.M. Dotson, A. Duarte, The importance of diversity in cognitive neuroscience, *Ann. N. Y. Acad. Sci.* 1464 (2020) 181–191, <https://doi.org/10.1111/nyas.14268>.
- [7] L. Jäncke, Sex/gender Differences in Cognition, Neurophysiology, and Neuroanatomy [version 1; Peer Review: 3 Approved, F1000Research, 7(F1000 Faculty Rev)], 2018, p. 805, <https://doi.org/10.12688/f1000research.13917.1>.
- [8] F.S. Ahmed, L.S. Miller, Executive function mechanisms of theory of mind, *J. Autism Dev. Disord.* 41 (5) (2011) 667–678, <https://doi.org/10.1007/s10803-010-1087-7>.
- [9] V. Ainley, L. Maister, M. Tsakiris, Heartfelt empathy? No association between interoceptive awareness, questionnaire measures of empathy, reading the mind in the eyes task or the director task, *Front. Psychol.* 6 (2015), <https://doi.org/10.3389/fpsyg.2015.00554>. Article 554.
- [10] M. Pardini, P.F. Nichelli, Age-related decline in mentalizing skills across adult life span, *Exp. Aging Res.* 35 (1) (2009) 98–106, <https://doi.org/10.1080/03610730802545259>.
- [11] S. Baron-Cohen, D.C. Bowen, R.J. Holt, C. Allison, B. Auyeung, M.V. Lombardo, P. Smith, M.C. Lai, The “Reading the Mind in the Eyes” Test: complete absence of typical sex difference in ~400 men and women with autism, *PloS One* 10 (8) (2015), e0136521, <https://doi.org/10.1371/journal.pone.0136521>.
- [12] P. Guariglia, L. Piccardi, F. Giaimo, S. Alaimo, G. Micciche, G. Antonucci, The eyes test is influenced more by artistic inclination and less by sex, *Front. Hum. Neurosci.* 9 (2015), <https://doi.org/10.3389/fnhum.2015.00292>. Article 292.
- [13] J.M. Moran, Lifespan development: the effects of typical aging on theory of mind, *Behav. Brain Res.* 237 (2013) 32–40, <https://doi.org/10.1016/j.bbr.2012.09.020>.
- [14] G. Ridinger, M. McBride, Money affects theory of mind differently by gender, *PloS One* 10 (12) (2015), e0143973, <https://doi.org/10.1371/journal.pone.0143973>.
- [15] M. Vellante, S. Baron-Cohen, M. Melis, M. Marrone, D.R. Petretto, C. Masala, A. Preti, The “Reading the Mind in the Eyes” test: systematic review of psychometric properties and a validation study in Italy, *Cognit. Neuropsychiatry* 18 (4) (2013) 326–354, <https://doi.org/10.1080/13546805.2012.721728>.
- [16] V. Warrier, K.L. Grasby, F. Uzefovsky, R. Toro, P. Smith, B. Chakrabarti, J. Khadake, E. Mawbey-Adamson, N. Litterman, J.J. Hottenga, G. Lubke, D. I. Boomsma, N.G. Martin, P.K. Hatemi, S.E. Medland, D.A. Hinds, T. Bourgeron, S. Baron-Cohen, Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition, *Mol. Psychiatr.* 23 (6) (2018) 1402–1409, <https://doi.org/10.1038/mp.2017.122>.
- [17] J. Kynast, M. Polyakova, E.M. Quinke, A. Hinz, A. Villringer, M.L. Schroeter, Age- and sex-specific standard scores for the reading the mind in the eyes test, *Front. Aging Neurosci.* 12 (2021), 607107, <https://doi.org/10.3389/fnagi.2020.607107>.
- [18] P.A. Bos, D. Hofman, E.J. Hermans, E.R. Montoya, S. Baron-Cohen, J. van Honk, Testosterone reduces functional connectivity during the “reading the mind in the eyes” test, *Psychoneuroendocrinology* 68 (2016) 194–201, <https://doi.org/10.1016/j.psyneuen.2016.03.006>.
- [19] E. Chapman, S. Baron-Cohen, B. Auyeung, R. Knickmeyer, K. Taylor, G. Hackett, Fetal testosterone and empathy: evidence from the empathy quotient (EQ) and the “reading the mind in the eyes” test, *Soc. Neurosci.* 1 (2) (2006) 135–148, <https://doi.org/10.1080/17470910600992239>.
- [20] A. Olsson, E. Kopsida, K. Sorjonen, I. Savic, Testosterone and estrogen impact social evaluations and vicarious emotions: a double-blind placebo-controlled study, *Emotion* 16 (4) (2016) 515–523, <https://doi.org/10.1037/a0039765>.
- [21] J. van Honk, D.J. Schutter, P.A. Bos, A.W. Kruijff, E.G. Lentjes, S. Baron-Cohen, Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio, *Proc. Natl. Acad. Sci. U.S.A.* 108 (8) (2011) 3448–3452, <https://doi.org/10.1073/pnas.1011891108>.
- [22] R.D. Brinton, J. Yao, F. Yin, W.J. Mack, E. Cadenas, Perimenopause as a neurological transition state, *Nat. Rev. Endocrinol.* 11 (7) (2015) 393–405, <https://doi.org/10.1038/nrendo.2015.82>.
- [23] E.B. Gold, The timing of the age at which natural menopause occurs, *Obstet. Gynecol. Clin. N. Am.* 38 (3) (2011) 425–440, <https://doi.org/10.1016/j.ogc.2011.05.002>.
- [24] M. Loeffler, C. Engel, P. Ahnert, D. Alfermann, K. Arelin, R. Baber, F. Beutner, H. Binder, E. Brähler, R. Burkhardt, U. Ceglarek, C. Enzenbach, M. Fuchs, H. Glaesmer, F. Girlich, A. Hagendorff, M. Häntzsch, U. Hegerl, S. Henger, T. Hensch, A. Hinz, V. Holzendorf, D. Husser, A. Kersting, A. Kiel, T. Kirsten, J. Kratzsch, K. Krohn, T. Luck, S. Melzer, J. Netto, M. Nüchter, M. Raschpichler, F. G. Rauscher, S.G. Riedel-Heller, C. Sander, M. Scholz, P. Schönknecht, M. L. Schroeter, J.C. Simon, R. Speer, J. Stäker, R. Stein, Y. Stöbel-Richter, M. Stumvoll, A. Tarnok, A. Teren, D. Teupser, F.S. Then, A. Tönjes, R. Treudler, A. Villringer, A. Weissgerber, P. Wiedemann, S. Zachariae, K. Wirkner, J. Thiery, The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany, *BMC Publ. Health* 15 (2015) 691, <https://doi.org/10.1186/s12889-015-1983-z>.
- [25] S. Bölte, Reading the Mind in the Eyes Test für Erwachsene (dt. Fassung) von S. Baron-Cohen, J.W. Goethe-Universität, Frankfurt/Main, Germany, 2005.
- [26] J.A. Durlak, How to select, calculate, and interpret effect sizes, *J. Pediatr. Psychol.* 34 (9) (2009) 917–928, <https://doi.org/10.1093/jpepsy/jsp004>.
- [27] D. Lakens, Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs, *Front. Psychol.* 4 (2013), <https://doi.org/10.3389/fpsyg.2013.00863> le 863.
- [28] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Routledge, Abingdon, UK, 1977.
- [29] S. Olderbak, O. Wilhelm, A. Hildebrandt, J. Quoidbach, Sex differences in facial emotion perception ability across the lifespan, *Cognit. Emot.* 33 (3) (2019) 579–588, <https://doi.org/10.1080/02699931.2018.1454403>.
- [30] S.R. Davis, I. Lambrinoudaki, M. Lumsden, G.D. Mishra, L. Pal, M. Rees, N. Santoro, T. Simoncini, Menopause, *Nat. Rev. Dis. Primers* 1 (2015), 15004, <https://doi.org/10.1038/nrdp.2015.4>.
- [31] J. Buvat, M. Maggi, A. Guay, L.O. Torres, Testosterone deficiency in men: systematic review and standard operating procedures for diagnosis and treatment, *J. Sex. Med.* 10 (1) (2013) 245–284, <https://doi.org/10.1111/j.1743-6109.2012.02783.x>.
- [32] N. Sumien, J.T. Cunningham, D.L. Davis, R. Engelland, O. Fadeyibi, G.E. Farmer, S. Mabry, P. Mensah-Kane, O.T.P. Trinh, P.H. Vann, E.N. Wilson, R.L. Cunningham, Neurodegenerative disease: roles for sex, hormones, and oxidative stress, *Endocrinology* 162 (11) (2021), bqab185, <https://doi.org/10.1210/endo/bqab185>.
- [33] A. Bandura, R.H. Walters, *Social Learning and Personality Development*, Holt Rinehart and Winston, New York, USA, 1969.
- [34] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, 1986.
- [35] J.R. Harris, Where is the children’s environment? A group socialization theory of development, *Psychol. Rev.* 102 (3) (1995) 458–489.
- [36] M. Melchers, C. Montag, M. Reuter, F.M. Spinath, E. Hahn, How heritable is empathy? Differential effects of measurement and subcomponents, *Motiv. Emot.* 40 (2016) 720–730.
- [37] S. Lee, E.P. Jacobsen, Y. Jia, B.E. Snitz, C.H. Chang, M. Ganguli, Reading the mind in the eyes: a population-based study of social cognition in older adults, *Am. J. Geriatr. Psychiatr.* 29 (7) (2021) 634–642, <https://doi.org/10.1016/j.jagp.2020.11.009>.
- [38] D. Dodel-Feder, K.J. Ressler, L.T. Germaine, Social cognition or social class and culture? On the interpretation of differences in social cognitive performance, *Psychol. Med.* 50 (1) (2020) 133–145, <https://doi.org/10.1017/S003329171800404X>.
- [39] S.D. Harlow, M. Gass, J.E. Hall, R. Lobo, P. Maki, R.W. Rebar, S. Sherman, P. M. Sluss, T.J. de Villiers, STRAW + 10 Collaborative Group, Executive summary of the stages of reproductive aging workshop +10: addressing the unfinished agenda of staging reproductive aging, *J. Clin. Endocrinol. Metab.* 97 (4) (2012) 1159–1168, <https://doi.org/10.1210/jc.2011-3362>.