Johansson Nolaker, E.; Murray, K.; Happé, F.G.; Charlton, R.A. Cognitive and affective associations with an ecologically valid test of theory of mind across the lifespan. **Neuropsychology** *in press.*

Cognitive and affective associations with an ecologically valid test of theory of mind across the lifespan

Emilie Johansson Nolaker [1], Kim Murray [2], Francesca Happé [2] and Rebecca A. Charlton [1]

[1] Department of Psychology, Goldsmiths University of London

[2] Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London

Running Head: Theory of Mind across the adult lifespan

Corresponding Author: Rebecca A. Charlton, Department of Psychology, Goldsmiths University of London, New Cross, London, SE14 6NW. Tel: +44 (0) 20 79I9 7870; Fax: +44 (0) 20 79I9 7870. Email:

r.charlton@gold.ac.uk

Conflict of Interest: The authors declare that they have no conflict of interest.

Abstract

Objectives: Many studies have demonstrated that theory of mind (ToM) ability declines with increasing age. Research has found that ToM-age associations are often mediated by other cognitive abilities particularly executive function. However, older adults rarely complain about real-world ToM difficulties. It has been suggested that older adults may perform better in real-world situations compared to experimental settings.

Methods: We examined performance on the Strange Stories Film Task (SSFT) which has been designed to assess ToM using naturalistic, video scenarios. Sixty adults aged between 17-95 years old completed the SSFT, inhibitory control (Stroop) and working memory (Letter-Number Sequencing) measures, the basic empathy scale (cognitive and affective empathy), and the broad autism phenotype questionnaire.

Results: ToM performance correlated significantly with age, whereas performance on a control task did not. Partial correlations and stepwise regression analyses demonstrated that performance on the three SSFT ToM measures was explained by a combination of executive function and empathy measures, with age explaining none of the variance.

Conclusions: Using a naturalistic test of ToM, performance was shown to decline with age for ToM but not control scenarios. Across the lifespan, the variance in ToM performance was explained by cognitive abilities and empathy but not age. Age alone may not influence ToM ability, but may be associated with age-related changes in cognition and social-cognition.

Public Significance

Older age is associated with poorer ability to understand the mental states of others, although older adults ready complain about the quality of their social interactions. Using a video task of real- world scenarios, we found that although poorer theory of mind ability was associated with older age, performance was explained by other cognitive abilities. Age-related cognitive difficulties rather than age alone may explain changes in understanding others in later life.

Introduction

Theory of mind (ToM) is the ability to to attribute mental states, for example to recognize that others may have beliefs or desires that differ from one's own (Flavell, 1999). Basic ToM develops in childhood, with explicit tests such as false belief tasks being passed around 4 years and implicit tasks (e.g. anticipatory looking based on attributed belief) in the second year of life.   More complex understanding of others' mental states (e.g. recognizing the difference between irony and lying) continues to develop through childhood (Dumontheil, Apperly, & Blakemore, 2010). ToM is thought to be stable in adulthood but it has seldom been examined across the adult lifespan, although a greater number of studies have examined ToM in later life (see Henry, Phillips, Ruffman, & Bailey, 2013 for a review). ToM underlies successful social interactions and decline in this ability may impact social functioning giving rise to social isolation and loneliness. Therefore ToM difficulties may be particularly important in clinical populations and among older adults where social isolation has been shown to have a significant impact on health and well-being (Adams, Sanders, & Auth, 2004; Hawkley, Thisted, Masi, & Cacioppo, 2010).

The aspects of ToM that are measured, often differ across the lifespan but have generally focused on understanding social interactions and specifically the ability to attribute mental states. In childhood ToM is often examined using (1st-order) false belief tasks (the understanding that others may hold a different belief to oneself) (Baron-Cohen, Leslie, & Frith, 1985). More complex attribution of false beliefs is measured using 2nd-order ('Jane thinks John thinks X') false belief tasks, often used with adolescents or adults (Happé, 1995). Other more advanced tests of mentalising have been developed for use with adult populations. These measures, including tasks such as the Strange Stories (Happé, 1994) or Frith-Happe Triangles (Castelli, Happe, Frith, & Frith, 2000), generally require interpretation of a protagonist's intention regarding another's mental state, e.g. irony, persuasion, deception, white lies,

double-bluff, etc. Task success is generally measured by accurate attribution of characters' intentions.

However, research has shown that some individuals with real-life ToM difficulties (e.g. individuals with

autism spectrum disorders), often "pass" ToM test questions, suggesting that some lab-based measures

are not sensitive to subtle real world difficulties (Scheeren, de Rosnay, Koot, & Begeer, 2013). The

reason for this may be related to the speed and complexity of real world social interactions, where

individuals must both interpret and react to events on a moment-by-moment basis (Scheeren et al.,

2013). The amount of Mental State Speech present in verbal responses has also been measured as an

indirect metric of ToM ability (Happé, 1994). Studies have demonstrated reduced use of Mental State

terms (despite correct interpretation of intention) in individuals with ToM difficulties (Happé, 1994).

More recently ToM tasks have also begun to examine the ability to generate appropriate behaviour

based on mental state attribution, e.g. when presented with an ambiguous scenario, "What would you

do/say in that situation?" It is hoped that such tasks may more accurately reflect real world behaviours

by measuring both understanding of the scenario and pro-social behaviours. Studies have identified

poorer pro-social responses in individuals reporting more autistic traits, suggesting that this may reflect

an additional aspect of ToM (Jameel, Vyas, Bellesi, Roberts, & Channon, 2014; Murray et al., 2017).

When ToM has been examined across the adult lifespan, objective measures show mixed

results. Duval et al. (2011) found no age effects from young to middle adulthood, but significantly

poorer performance in older adults on an attribution of intention task and 2nd-order false belief tasks.

They further found that age had a direct effect on 2nd-order false belief performance, whereas the

association between age and 1st-order false belief performance was mediated by executive function. In

contrast Bernstein et al. (2011), using a continuous performance false belief task found greater false

belief bias in middle-aged and older adults compared to young adults, as well as a general reduction in

ToM performance from young to old age. In a subsequent regression analysis, executive function, working memory, episodic memory and processing speed all failed to explain any additional variance in false belief performance beyond age effects. Although these studies used different measures and analyses, the contrast in results is still striking, with one study demonstrating middle-aged adults perform like young adults (Duval, Piolino, Bejanin, Eustache, & Desgranges, 2011) and the other demonstrating they perform like older adults (Bernstein, Thornton, & Sommerville, 2011); in one study age-ToM associations are affected by cognitive abilities (Duval et al., 2011) and in the other they are independent (Bernstein et al., 2011). There are notable differences in the ages of the groups used in these studies (mean age difference between young-middle groups, Duval et al. = 29 years, Bernstein et al. = 37 years; mean age difference between middle-old groups, Duval et al. = 18 years, Bernstein et al. = 11 years) and many analyses used age-group rather than age as a continuous variable. Age differences and arbitrary age groupings may have influenced the different patterns of results by emphasising group differences without considering the continuous nature of age.

More consistent results have been observed examining ToM performance in later life. Many studies have demonstrated poorer ToM abilities in older compared to younger adults even when accounting for intelligence or other cognitive abilities (Charlton, Barrick, Markus, & Morris, 2009; Rakoczy, Harder-Kasten, & Sturm, 2012). Furthermore a meta-analysis including a range of different ToM measures, supported poorer performance in later-life and a moderate effect size across all measures (Henry et al., 2013). It is worth noting that several studies have found equivalent performance between older and younger adults (Saltzman, Strauss, Hunter, & Archibald, 2000; Keightley, Winocur, Burianova, Hongwanishkul, & Grady, 2006). Only one paper has demonstrated superior performance among older versus young adults (Happé, Winner, & Brownell, 1998); the authors note, however, that

the young and old groups were not IQ-matched, and it seems likely that selection effects explain this result. Many studies have demonstrated that age-ToM associations are partly or fully mediated by other cognitive abilities, with some variability in results that may be due to the specific cognitive ability (Charlton et al., 2009; Maylor, Moulson, Muncer, & Taylor, 2002; Rakoczy et al., 2012). Working memory and inhibitory control have been described as core deficits in ageing (Charlton, Barrick, Lawes, Markus, & Morris, 2010; Hasher & Zacks, 1988), and previous studies have suggested that these abilities may influence ToM performance (Charlton et al., 2009; Maylor et al., 2002; Rakoczy et al., 2012; Henry et al., 2013). Furthermore, the task demands of many ToM tests require working memory abilities (holding information in mind, evaluating it and then generating a response), and inhibitory control (inhibiting responses based on own knowledge in order to answer test questions about a character's false belief). In a typical ageing population, seven (out of eight) executive function tests correlated significantly with ToM performance, including domains of working memory and inhibitory control (Charlton et al., 2009). Robust correlations between ToM and executive function abilities have also been noted both in age-related disease such as stroke (Pluta, Gawron, Sobanska, Wojcik, & Lojek, 2017). Research suggests an important role for executive function abilities in ToM performance across the adult lifespan, but it is not clear whether ToM performance is affected by age directly or by age-related changes in other cognitive domains. Understanding the mechanism for age-related changes in ToM abilities will allow us to intervene to reduce social problems and social isolation that has a significant negative impact in older age (Bailey, Henry, & Von, 2008). To date few studies have directly examined the influence of other socio-cognitive abilities such as emotional empathy on ToM performance in ageing.

In recent years the distinction has been made between 'cognitive' and 'affective' ToM (Duval et al., 2011; Mitchell & Phillips, 2015), that is between attribution of mental states and affective empathy

for others' feelings. Cognitive ToM has been described as the cognitive understanding of the beliefs and intentions of others, whereas affective ToM has been described as requiring empathic understanding of others' emotions (Mitchell & Phillips, 2015). These definitions often overlap with descriptions of cognitive and affective empathy. Empathy is often described as an emotional response relating to someone else's affective state, with cognitive empathy being the representation of others' mental states (i.e. ToM) and affective empathy being the recognition of and response to others' emotions (Blair, 2005). Although different aspects of empathy are often strongly associated with each other, they may also dissociate particularly in clinical populations including autism spectrum disorders (Sucksmith, Allison, Baron-Cohen, Chakrabarti, & Hoekstra, 2013). For example, although studies have identified difficulty with cognitive empathy/ToM in individuals with autism spectrum disorders, many have found no difficulties in affective empathy (Rogers, Dziobek, Hassenstab, Wolf, & Convit, 2007). Furthermore ToM and empathy have been shown to dissociate in other neurodevelopmental groups (e.g. Conduct disorder with callous/unemotional traits versus ASD; Jones, Happé, Gilbert, Burnett, & Viding, 2010). The presence of neurodevelopmental traits (e.g. the broad autism phenotype, BAP) is significantly associated with emotional empathy abilities in relatives of individuals with autism spectrum disorders and the general population (Grove, Baillie, Allison, Baron-Cohen, & Hoekstra, 2014; Lamport & Turner, 2014). However, few studies have examined relationships between ToM, affective empathy and traits associated with the BAP in ageing (Bailey et al., 2008; Mitchell & Phillips, 2015). To our knowledge one study has examined the relationship between ToM and cognitive versus affective empathy. Bailey et al. (2008) demonstrated that older adults performed significantly worse than young adults on an affective ToM task (Reading the Mind in the Eyes Task; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), as well as reporting poorer cognitive empathy on a self-report questionnaire. However, no age differences were observed on self-reported affective empathy. In other studies of cognitive versus affective abilities in old age, findings are mixed. Bottirolli et al. (2016) demonstrated poorer performance for older adults

(compared to young adults) on a cognitive but not affective ToM measure (a Faux Pas task designed to include both cognitive and affective measures). However other studies have also identified age-effects on affective ToM measures (Mahy et al., 2014; Duval et al., 2011). Overall, these results suggest there may be a different pattern of age-related decline in cognitive versus affective abilities.

An important consideration when discussing ToM across the neurotypical lifespan, is that any differences or difficulties in performance are likely to be small in comparison to clinical disorders associated with ToM difficulties (i.e. Autism Spectrum Disorders; or Schizophrenia). Indeed, even among clinical groups, individuals may perform well on lab-based tests despite difficulties in the real world (Murray et al., 2017). In order to detect the subtle differences in performance that are likely to occur across the lifespan we require measures that are both sensitive and ecologically valid. It has been suggested that video-based tasks may overcome some of the difficulties with traditional lab-based tasks (Heavey, Phillips, Baron-Cohen, & Rutter, 2000; Murray et al., 2017; Sullivan & Ruffman, 2004). The advantage of videos is that they allow the dynamic presentation of complex scenes with multiple sources of information that may be both subtle and fleeting. In this study we utilise a recently developed video ToM task (the Strange Stories Film Task, SSFT) designed to assess ToM in a naturalistic manner in adults (Murray et al., 2017). The SSFT was developed based on the Strange Stories Task (Happé, 1994), as a series of video scenarios showing the same two characters and requiring interpretation of speaker's intention and generation of an appropriate response (see Methods section for a more detailed description; full details of development can be found in Murray et al. (2017)). The SSFT includes assessment of several different aspects of ToM; i) correct identification of the *intention* of a character, ii) measurement of use of *mental state speech*, and iii) correct generation of a suitable *social interaction*. Understanding others' mental states, such as intentions is the core feature of ToM according to most

definitions (Mitchell & Phillips, 2015), and mental state speech has been used as an index of a participants ability to think about others' mental states (Happé, 1994). Less widely used are questions requiring a response describing an interaction, where a participant must generate their "own" social response to each scenario. A previous study has shown that individuals reporting high autistic traits responded less pro-socially to scenarios (Jameel et al., 2014), thus this measure may be a useful reflection of real world behaviour. By using a test that mimics real world social situations, measuring multiple aspects of ToM we hope to better understand the subtle changes in ToM across the lifespan.

Given that "cognitive empathy" has been shown to mediate age-social functioning associations (Bailey et al., 2008) and the known negative consequences of social isolation in ageing (Blozik et al., 2009), understanding these associations may inform well-being in older age. In this study we examined ToM performance cross-sectionally across the adult lifespan using a naturalistic measure. We further examined associations between ToM performance with age, executive function abilities (specifically working memory and inhibitory control), cognitive and affective empathy. We hypothesise that ToM performance will decline with increasing age, but that this will be explained by age-related declines in domain-general cognitive abilities. We hypothesize that ToM performance will be associated with self-rated cognitive but not affective empathy.

Methods

Participants

Sixty adults aged 17-95 (M=44, SD=18.5) were recruited to participate in the study. One participant (aged over sixty) was excluded from the analyses, having had a stroke that significantly influenced his

cognitive ability. On the Stroop task (see below) he performed more than two standard deviations below the mean of his age group, and was excluded from the analysis. Demographic data on the remaining 59 participants are presented in Table 1 where the sample has been divided into three age groups for descriptive purposes, although age is used as a continuous variable in all analysis. All participants provided written informed consent. After the study, all participants were debriefed both verbally and via a written debrief form. The study was approved by Goldsmiths University of London Ethics Committee.

Younger participants (aged 17-34 years) were students from Goldsmiths University. Twenty students were from the Department of Psychology and were assigned student research participation credits. Middle-aged (n=20; aged 35-59 years) participants were recruited at random from a cafe in South-east London. Participants were given a hot drink as a thank you for their participation. Older participants (aged 60-95 years) were recruited from an elderly independent housing community in the London Borough of Camden. The community offers housing opportunities for older adults with a low income, where people live independently in an apartment within a block for older adults. Participants aged over 60 years old were screened for risk for dementia using the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975). All older participants scored above the cut-off of 24 ruling out likely dementia (Mean=28.53, SD=1.3; Tombaugh & McIntyre, 1992). All participants were fluent in English, any non-native speaker had lived in the UK for at least 10 years. None of the participants reported problems with hearing and all had corrected-to-normal or normal vision. Participants were tested in a quiet environment in their own homes, in a quiet room in the café or in a laboratory at Goldsmiths University. All participants were tested independently in the company of the experimenter.

Materials

Strange Stories Film Task. The Strange Stories Film task (SSFT; Murray et al., 2017) was used to assess ToM and social cognition in a naturalistic manner. The SSFT was developed and tested on 40 adults (20 with autism spectrum disorders and 20 without). The SSFT comprised 15 short video clips acted by two semi-professional actors, using similar social scenarios as Happé's Strange Stories (Happé, 1994). Twelve scenarios involved mental states such as telling a white lie, using persuasion, or double bluff (henceforth referred to as ToM scenarios), and three were non-ToM scenarios (henceforth referred to as Control scenarios). The actors spoke directly to the camera, to mimic real world conversations from a first-person perspective. After watching each video clip participants were asked three questions examining *Intention* (i.e., Why did the character say that?), *Interaction* (ie If you were in the other character´s situation, what would you say next?), and a memory question to check for attention or gross memory difficulties (e.g., Where was the character?). Participants answered verbally and their responses were transcribed by the researcher. Intention questions were scored between 0-2 depending on accuracy and appropriateness; these responses were also coded for presence of *mental state* and metacognitive speech (0-2 points, henceforth "Mental State Speech"). Interaction questions were scored between 0-2 depending on appropriateness. The memory question was scored 0-1 for accuracy. The Intention question was designed to measure ToM, whereas the Interaction and Mental State Speech responses were intended to measure social cognition more broadly. All scoring followed guidelines in Murray et al (2017) and was performed by a single rater (EJN). To assess consistency of scoring, inter-rater reliability was calculated with an experienced coder (RAC) and two other research assistants in the lab, using two-way random model intra-class correlations (absolute values) in IBM SPSS Statistics, version 22 (IBM Corp., 2013), all values were deemed good and were as follows: ToM- Intention, $r=.962$; ToM-Mental State Speech, $r=.871$; ToM-Interaction, $r=.916$; ToM-Memory, $r=.969$; Control-Intention, $r=.894$; Control -Mental State Speech, $r=.907$; Control –Interaction, $r=.724$; Control –Memory, no variance was observed between raters.

The sensitivity of the SSFT has been reported previously in young adults and adults with ASD (Murray et al., 2017), but has not been widely used or used across the lifespan. For reference the reliability of the measure is reported here. For the ToM scenarios Cronbach's alphas were as follows: Intention (alpha =.57); Mental State Speech (alpha =.48); and Interaction (alpha =.55). For Memory questions there was very little variance in the data and Cronbach's alphas were not deemed appropriate: for young and midlife adults most individuals scored at ceiling (maximum score=12, achieved by: young=19/20; midlife=18/20); for older adults results had only slightly more variance (8 individuals scored=12; 7 scored=11; 3 scored=10; 1 scored=9). For the Control scenarios Cronbach's alphas were robust for the Interaction scores (=.77), but much weaker for Mental State Speech (alpha =-.001) and Intention (alpha =-.030); note however, that these variables were not intended to be meaningful for the control clips. Memory scores were at ceiling, with all but three individuals (two midlife, one older) achieving the maximum score. Scores from the current data are reported in Table 2.

Cognitive and Empathy Measures. Inhibitory control was measured by the Stroop test (Golden, 1978); completion time on the Colour-Ink subscale minus completion time for the Colour-Word subscale was used in the analysis therefore lower scores indicate better performance. Working memory was measured by the Letter-Number Sequencing task from the Wechsler's Intelligence Test IV (Wechsler, 2008); the total raw score was used in the analysis with higher scores indicating better performance. Raw scores (rather than age-corrected scores) were used for both cognitive measures in order to allow examination of age-effects. The Basic Empathy Scale (BES) was used to measure cognitive and affective empathy (Jolliffe & Farrington, 2006), with questions such as "I'm not usually aware of my friend's feelings", and "I tend to feel scared when I'm with friends who are afraid" respectively. The Broad

Autism Phenotype Questionnaire (BAPQ; Hurley, Losh, Parlier, Reznick, & Piven, 2007) was used to assess presence of autism traits. As traits are considered to vary across the population, participants with high scores were not excluded. For reference, six individuals scored above the suggested BAPQ cut-off (>3.35 for males and >3.25 for females; Hurley et al., 2007), three men (aged 37, 65, 95) and three women (aged 19, 20, 46), although it is worth noting that these cut-offs were developed in a sample of young and mid-life adults.

Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics, version 22 (IBM Corp., 2013). The relationship between age and ToM performance was investigated using Pearson's product moment correlations. The associations between ToM performance, age, executive function abilities, cognitive and affective empathy were examined using Pearson's product moment correlations. Stepwise linear regressions were used to examine which variables best explained the variance in each aspect of ToM ability measured by the SSFT. To assure that gender did not influence results, gender differences on SSFT measures were calculated using independent sample t-test. Multiple comparison correction was performed using the Holm-Bonferroni sequence correction.

Results

Age effects. For descriptive purposes, the means and standard deviations are presented by age-group in Table 2. Mean scores for cognitive measures are reported in Table 1, no age-group differences are noted for the working memory (Letter-Number Sequencing) age-scaled scores but the non-age corrected difference between the Stroop Colour and Interference tasks demonstrate expected age

differences (slower performance with older age). Please note that age is used as a continuous variable in all analysis presented below.

Gender differences. No gender differences were noted in highest education level, but were apparent in age, see Table 3. For the SSFT, gender differences were observed on ToM Intention and Mental State Speech , but no other measures. After correction for multiple comparisons, only ToM Intention ability was demonstrated significant gender differences. See Table 3 for full details. Women scored more highly on affective empathy (Women: mean=42.06, sd=6.90; Men: mean=37.74, sd=6.96; F=5.46, p=.023); but not cognitive empathy (Women: mean=38.06, sd=3.45; Men: mean=37.22, sd=4.83; F=.545, p=.464).

Correlation Analyses.

Age. Significant negative correlations were observed between age and each aspect of ToM (see Figure 1 and Table 4), indicating declining performance as age increases. All associations remained significant after controlling for multiple comparison correction. For Control SSFT scores, age correlated significantly with Mental State Speech and Interaction scores, but not with Intention or Memory scores; results did not survive multiple comparison correction.

Age correlated significantly with performance on both inhibitory control (r=.729, p<.001) and working memory (r=-.373, p=.004). No significant correlations were observed between age and either affective (r=-.135, p=.310) or cognitive (r=-.039, p=.770) empathy, or BAP traits (r=-.176, p=.181).

BAP Traits. Self-reported BAP traits did not correlate significantly with any measure of ToM. However, partial correlations between BAP traits and ToM performance controlling for age, reveal a significant association between BAP traits and ToM Intention ability (see Table 4). Results did not survive multiple comparison correction.

Cognitive function. All aspects of ToM performance (Intention, Mental State Speech, Interaction) correlated significantly with both inhibitory control and working memory measures, see Table 5. Results indicate better ToM performance was associated with better cognitive abilities. If age was controlled for in a partial correlational analysis the association between ToM performance and working memory remained significant.

Empathy. ToM Intention ability correlated significantly with both cognitive and affective empathy, see Table 5. Mental State Speech correlated significantly with cognitive empathy but not affective empathy. No significant association was observed between ToM Interaction ability and either cognitive or affective empathy scores. The pattern of results remained the same if a partial correlation was performed controlling for age, see Table 5. Also see Table 5 for details of multiple comparison correction. Note, the correlation between cognitive and affective empathy was not significant (r=.215, p=.102), suggesting that these subscales measure different constructs.

Regression Analysis. Forward stepwise regression analyses were performed with each aspect of ToM performance (Intention, Mental State Speech, Interaction) in turn as the dependent variable.

Included as independent variables were age, education, gender, working memory, inhibitory control, cognitive and affective empathy, and BAP traits. Full results are presented in Table 6.

For ToM Intention, working memory, affective empathy, control and gender contributed to explaining 49.9% of the variance in performance. Age did not contribute to the model. For Mental State Speech, working memory, gender and cognitive empathy contributed, explaining 41.7% of the variance in performance. Variance in ToM Interaction ability (37.9%) was explained by performance in inhibitory control and education level. Memory performance on the ToM SSFT task (37.4%) was explained by age and working memory ability. See Table 5 for full details and statistics.

Discussion

The aim of this study was to examine ToM abilities across the lifespan using a naturalistic video-based test, and to examine associations with executive function, cognitive and affective empathy. In keeping with previous studies in ageing (Bottiroli, Cavallini, Ceccato, Vecchi, & Lecce, 2016; Henry et al., 2013), all three ToM measures derived from the SSFT (and the memory question) correlated negatively with age. Despite using a measure designed to optimise ecologically validity, age-related declines in performance were still observed. In previous studies examining ToM across the lifespan, analyses have largely explored group differences between young, middle-aged and older adults, and have found age effects either before (Young>Mid-Iife=Old) or after (Young=Mid-Iife> Old) middle-age (Duval et al., 2011; Bernstein et al., 2011). In contrast, this study considered age as a continuous variable and significant correlations were noted between the three aspects of ToM measured and age (see Figure 1). No correlations with age were observed for either cognitive or affective empathy, suggesting that self-reported empathy is not associated with age. This finding is in keeping with a previous study which examined cognitive and affective ToM, which found no age-associations on self-reported cognitive or

affective measures (Duval et al., 2011). However another study described older adults as self-reporting lower scores on cognitive empathy compared to young adults, but no age-group differences on self-report affective empathy (Bailey et al., 2008).

Some authors consider cognitive empathy and ToM to be synonymous (Mitchell & Phillips, 2015) but if they are then there are discrepancies in the age effects observed in this study, with age effects observed on "objective" task performance (SSFT performance) but not "subjective" self-report measures (BES). In the current study, on the objective measure, ToM Intention and Mental State Speech did correlate significantly with the subjective self-report measure of cognitive empathy, but only modestly (r≈.3) and no correlation was observed between self-report cognitive empathy and ToM Interaction ability. Furthermore subjectively reported affective empathy significantly correlated with objectively measured ToM Intention ability, but not the other two ToM measures (Mental State Speech or Interaction). Similar results to those found here were identified in a previous study, where no significant correlations were noted between objective and subjective ToM measures (Duval et al., 2011). These results may suggest that perhaps cognitive empathy and ToM do not measure the same underlying construct, although they may share key components (Mitchell & Phillips, 2015). Alternatively, results could reflect discrepancies between objective tasks performance versus subjective self-report measures (usually questionnaires), where individuals may under-estimate their difficulties, or perhaps objective lab-based measures are more strongly influenced by cognitive demands than everyday life social skills, the latter reflected in better self-reported abilities.

Interestingly, this study did not identify significant correlations between the three aspects of ToM measured and BAP traits, although once age-effects were accounted for ToM Intention correlated

significantly with BAP traits. This finding is very similar to that of Murray et al. (2017) who found that

ToM Intention performance correlated significantly with Autism Quotient scores in typical young adults,

although the association did not reach the high significance level imposed (p<.01). These findings may

reflect the fact that although ToM difficulties reflect the core social communication feature in ASD, they

are not the only important diagnostic characteristic.

The results of both the correlation and regression analyses suggest that executive function and

empathy explain a proportion of the variance in ToM performance, whereas age does not contribute

statistically to the regression models. Results suggests that working memory contributes to ToM

performance across the lifespan, not just in ageing. However, the association between ToM

performance and inhibitory control no longer reached significance after controlling for age. Although

age correlated significantly with ToM abilities, it was not the most important metric for explaining

performance. ToM Intention was explained by working memory, affective empathy and gender; and

ToM Interactions by inhibitory control and education. Variance in Mental State Speech was explained by

working memory, gender and cognitive empathy. Age did not contribute to explaining the variance on

any of the measure of ToM. Only for the memory component of ToM did age explain a large proportion

of the variance in performance. Although some previous studies have found that age-ToM associations

remain after taking into account executive function (Bernstein et al., 2011; Bottiroli et al., 2016; Maylor

et al., 2002), others have found that executive function mediates age-ToM associations (Bailey & Henry,

2008; Charlton et al., 2009; Duval et al., 2011; Rakoczy et al., 2012). Results from this study support the

importance of executive function abilities for ToM performance, not only in ageing but across the adult

lifespan. Few studies have explicitly assessed the impact of both executive function and empathy on

ToM across the lifespan (Duval et al., 2011). The results of this study suggest that cognition and empathy

both contribute to ToM performance, and the results of the partial correlations controlling for age

suggest that these abilities are important not just in ageing but across the adult lifespan.

Results of this study should be considered keeping in mind some limitations. Recruitment

occurred in different settings depending on age (young adults were predominantly university students,

whereas middle-aged adults were recruited at a café, and older adults at an independent residential

association), and this may lead to sampling biases. As in many ageing studies, the older individuals who

participated were relatively healthy, with 8 taking 1-2 medications and only 1 taking three or more

medications. This compares to UK national statistics showing 88% of those aged over 65 take at least

one medication, and 68% take at least three medications (Scholes, Faulding, & Mindell, 2014). This

suggests that this community housing sample are not unusually unwell for their age. Age differences

were also noted between men and women. We focused on two aspects of executive function (inhibitory

control and working memory), but future studies may benefit from examining the contribution of other

aspects of executive function and a formal measure of episodic memory. In addition, this study used a

recently developed, novel ToM task. Although our study demonstrates a pattern of results similar to

studies using more established measures and a previous study has demonstrated convergent validity for

ToM SSFT Intention scores compared to the widely used Strange Stories task (Murray et al., 2017), we

did not include a comparison ToM measure in this study. In future studies it would be useful to compare

the SSFT to established ToM measures. Another potential limitation of the SSFT task across the lifespan

is the use of young actors in the scenarios. Previous research has demonstrated own-age biases in

recognition memory tasks, but the effect of own-age biases in attention is less clear (Harrison & Hole,

2009). Future task development could expand the scenarios to include actors across a wider age-range,

to allow the examination of possible own-age biases. Furthermore, as the SSFT was designed as an

ecologically valid measure, future studies could assess real-life ToM abilities through self and other-report, to examine whether the SSFT is providing information valid to the real-world.

In summary, in this lifespan study of ToM we found that even using a test designed to have ecological validity, performance on ToM measures was negatively associated with age. However ToM performance was largely explained by executive function and empathy, rather than age. Results suggest that age-related ToM changes may be due to other cognitive and socio-cognitive changes rather than age alone. This suggests that declines in ToM performance, associated difficulties in social interaction and possible sequelae of social isolation, may not be inevitable in ageing. By supporting cognitive functions it may be possible to reduce the impact of ToM changes in later life. Further studies are required to fully understand the associations between socio-cognitive and social functions across the lifespan, in order to mediate the potentially negative consequences on well-being.

References

Adams, K. B., Sanders, S., & Auth, E. A. (2004). Loneliness and depression in independent living retirement communities: risk and resilience factors. *Aging & Mental Health, 8,* 475-485.

Bailey, P. E., Henry, J. D., & Von, H. W. (2008). Empathy and social functioning in late adulthood. *Aging & Mental Health.12(4):499-503*.

Bailey, P. E. & Henry, J. D. (2008). Growing Less Empathic With Age: Disinhibition of the Self-Perspective. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 63,* 219-226.

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

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A study with normal adults, and adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 241-251.

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of Mind Through the Ages: Older and Middle-Aged Adults Exhibit More Errors Than Do Younger Adults on a Continuous False Belief Task. *Experimental Aging Research, 37,* 481-502.

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition, 14,* 698-718.

Blozik, E., Wagner, J. T., Gillmann, G., Iliffe, S., von Renteln-Kruse, W., Lubben, J. et al. (2009). Social network assessment in community-dwelling older persons: results from a study of three European populations. *Aging-Clinical & Experimental Research.21(2):150-7*.

Bottiroli, S., Cavallini, E., Ceccato, I., Vecchi, T., & Lecce, S. (2016). Theory of Mind in aging: Comparing cognitive and affective components in the faux pas test. *Archives of Gerontology and Geriatrics, 62,* 152-162.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12,* 314-325.

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 Ageing. *Psychology & Aging, 24,* 338-348.

Charlton, R. A., Barrick, T. R., Lawes, I. N. C., Markus, H. S., & Morris, R. G. (2010). White matter pathways associated with working memory in normal aging. *Cortex, 46,* 474-489.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13,* 331-338.

Duval, C., Piolino, P., Bejanin, A., Eustache, F., & Desgranges, B. (2011). Age effects on different components of theory of mind. *Consciousness and Cognition, 20,* 627-642.

Flavell, J. H. (1999). Cognitive development: children's knowledge about the mind. *Annual Review of Psychology., 50,* 21-45.

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

Golden, C. J. (1978). Stroop Color and Word Test: A Manual for Clinical and Experimental Uses. Chicago, IL, Skoelting.

Grove, R., Baillie, A., Allison, C., Baron-Cohen, S., & Hoekstra, R. A. (2014). The latent structure of cognitive and emotional empathy in individuals with autism, first-degree relatives and typical individuals. *Molecular Autism, 5,* 42.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24,* 129-154.

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66,* 843-855.

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

Hasher, L. & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H.Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol 22 ed., pp. 193-225). San Diego, USA: Academic Press, Inc.

Hawkley, L. C., Thisted, R. A., Masi, C. M., & Cacioppo, J. T. (2010). Loneliness Predicts Increased Blood Pressure: 5-Year Cross-Lagged Analyses in Middle-Aged and Older Adults. *Psychology & Aging, 25,* 132-141.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: a naturalistic measure of social understanding in autism. *Journal of Autism & Developmental Disorders, 30,* 225-236.

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,* 826-839.

IBM Corp. (2013). IBM SPSS Statistics for Windows, version 22.0. Armonk, NY, IBM Corp.

Jameel, L., Vyas, K., Bellesi, G., Roberts, V., & Channon, S. (2014). Going "Above and Beyond": Are Those High in Autistic Traits Less Pro-social? *Journal of Autism and Developmental Disorders, 44,* 1846-1858.

Jolliffe, D. & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence, 29,* 589-611.

Jones, A. P., Happé, F. G. E., Gilbert, F., Burnett, S., & Viding, E. (2010). Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 51,* 1188-1197.

Keightley, M. L., Winocur, G., Burianova, H., Hongwanishkul, D., & Grady, C. L. (2006). Age Effects on Social Cognition: Faces Tell a Different Story. *Psychology and Aging, 21,* 558-572.

Lamport, D. & Turner, L. A. (2014). Romantic Attachment, Empathy, and the Broader Autism Phenotype among College Students. *The Journal of Genetic Psychology, 175,* 202-213.

Mahy, C. E. V., Vetter, N., K++hn-Popp, N., L+Âcher, C., Krautschuk, S., & Kliegel, M. (2014). The influence of inhibitory processes on affective theory of mind in young and old adults. *Aging, Neuropsychology, and Cognition, 21,* 129-145.

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

Mitchell, R. L. C. & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia, 70,* 1-10.

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N. et al. A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Research,* (in press).

Pluta, A., Gawron, N., Sobanska, M., Wojcik, A. D., & Lojek, E. The Nature of the Relationship Between Neurocognition and Theory of Mind Impairments in Stroke Patients. *Neuropsychology,* (in press).

Rakoczy, H., Harder-Kasten, A., & Sturm, L. (2012). The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology, 103,* 58-72.

Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O. T., & Convit, A. (2007). Who Cares? Revisiting Empathy in Asperger Syndrome. *Journal of Autism and Developmental Disorders, 37,* 709-715.

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

Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 54,* 628-635.

Scholes, S., Faulding, S., & Mindell, J. (2014). Prescibed medicines. In R.Craig & J. Mindell (Eds.), *Health Survey for England, 2013* (Leeds, UK: Health and Social Care Information Centre.

Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia, 51,* 98-105.

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

Tombaugh, T. N. & McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society, 40,* 922-935.

Wechsler, D. (2008). Wechsler Adult Intelligence Scale - IV. 4th edition. San Antonio, USA, Psychological Corporation.

Figure Headings

Figure 1. Scatterplots showing correlations between age and SSFT Intention, Mental State Speech and Interaction scores.
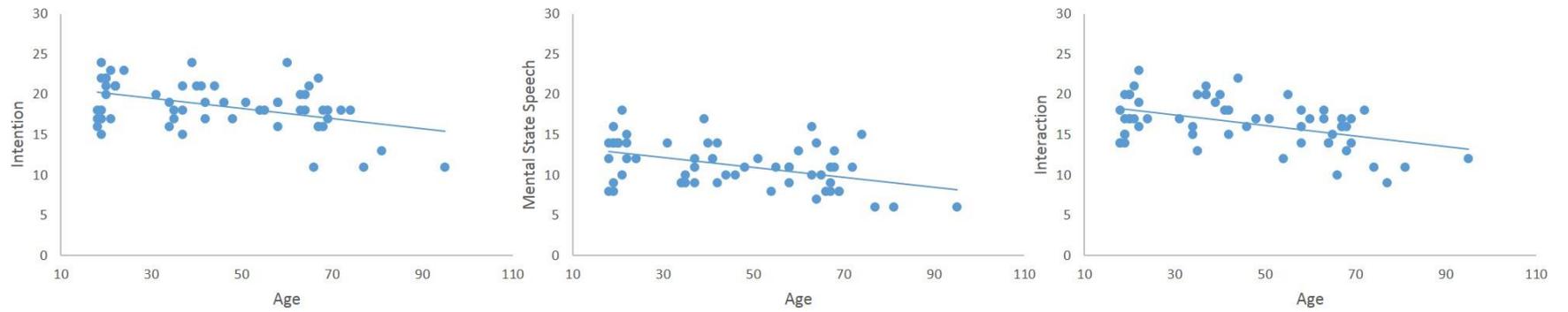
Table 1: Means (and standard deviations) for demographic data divided by age-group for descriptive purposes only.

	17-34 years (N=20)	35-59 years (N=20)	60-95 years (N=19)
Age	21 (4.17)	44 (8.40)	70 (8.0)
Gender (Women:Men)	17:3	9:11	10:10
Education [a]	3.35 (0.58)	4.30 (0.98)	2.84 (1.30)
Mental Health [b]	0 (0)	0.25 (0.64)	0.37 (0.68)
Medication frequency [c]	No medication use	2 people taking 1-2 medications	8 people taking 1-2 medications; 1 person taking three or more medications
BAPQ [d]	99.29 (18.52)	95.50 (22.33)	88.05 (25.49)
MMSE [e]	-	-	28.53 (1.30) Range 26-30
Letter-Number Sequencing Scaled Score [f]	10.45 (2.31)	11.40 (2.66)	10.95 (1.99)
Stroop Difference in Seconds (Interference-Colour) [g]	45 (14)	67 (18)	109 (34)

[a] Education level: 0 = Left school, 1 = Vocational qualification including apprenticeships, 2 = General Certificate of Secondary Education (GCSE, UK exams at 16 years old), 3 = Advanced-level (A-level, UK exams at 18 years old, 4 = Undergraduate Degree, 5 = Postgraduate Degree, 6 = PhD or higher

[b] Mental health problems: 0= Never, 1= One occasion, 2= Two-three occasions, 3= More than three occasions

[c] Current medication: 0=None, 1= One/Two, 2= Three or more

[d] BAPQ= Score on the Broad Autism Phenotype Questionnaire (Max: 216)

[e] MMSE, Mini Mental State Examination, only performed for adults aged over 60 years old

[f] $F=.825$, $p=.444$

[g] $F=39.1$, $p<.001$

Table 2: Means and standard deviations for the SSFT scores, illustrated in three age groups.

	17-34 years (Young, n=20)	35-59 years (Middle-aged, n=20)	60-95 years (Older, n=19)	Whole sample (n=59)	Control group means (Mean age=30.6, n=20) from Murray et al., 2017
ToM					
Intention (Max: 24)	19.70 (2.72)	18.80 (2.07)	17.16 (3.66)	18.58 (3.01)	18.80 (2.33)
Mental State Talk (Max: 24)	12.75 (2.72)	10.95 (2.16)	10.00 (3.09)	11.25 (2.87)	13.75 (2.45)
Interaction (Max: 24)	17.35 (2.50)	17.55 (2.78)	14.47 (2.80)	16.49 (3.00)	16.95 (4.14)
Memory (Max: 12)	11.95 (0.22)	11.90 (0.31)	11.16 (0.90)	11.68 (0.66)	11.85 (0.37)
Control					
Intention (Max: 6)	4.15 (0.81)	4.85 (0.93)	4.53 (1.02)	4.51 (0.95)	4.40 (0.99)
Mental State Talk (Max: 6)	1.20 (1.01)	1.15 (1.04)	0.74 (0.65)	1.03 (0.93)	1.25 (1.16)
Interaction (Max: 6)	5.05 (1.39)	5.35 (0.99)	4.21 (1.27)	4.88 (1.30)	5.70 (0.65)
Memory (Max: 3)	3.00 (0)	2.90 (0.31)	2.95 (0.23)	2.95 (0.22)	3.00 (0)

Table 3: Means and standard deviations by gender, and group differences.

	Males	Females	Group differences
Age	53.48, (sd=17.7)	39.06 (sd=21.1)	t=-2.72, p=.009 $^\pm$
Highest Education level	3.52 (1.34)	3.50 (1.03)	t=-.070, p=.944
ToM SSFT scores			
Intention	17.09, (sd=2.9)	19.53 (sd=2.7)	**t=3.28 p=.002 $^\pm$**
Mental State Talk	10.17, (sd=2.4)	11.94 (sd=2.9),	t=2.41, p=.019 *
Interaction	15.83, (sd=3.4)	16.92 (sd=2.7)	t=1.37p=.175
Memory	11.48, (sd=.85)	11.81 (sd=.47)	t=1.70, p=.100
Control SSFT scores			
Intention	4.65, (sd=.98)	4.42 (sd=.94)	t=-.914, p=.365
Mental State Talk	1.0, (sd=.85)	1.06 (sd=.98)	t=.222, p=.825
Interaction	4.65, (sd=1.4)	5.03 (sd=1.2)	t=1.08, p=.283
Memory	2.96, (sd=.219)	2.94 (sd=.23)	t=-.202, p=.840

* p<.05, $^\pm$ p<.001; Bold = remains significant after multiple comparison testing.

Table 4: Correlations between SSFT scores and age and BAPQ scores.

	Correlations with Age	Correlations with BAPQ	Correlations with BAPQ controlling for Age
ToM SSFT scores			
Intention	**r=-.441, p<.001 ±**	r=-.210, p=.111	r=-.325, p=.013
Mental State Talk	**r=-.452, p<.001 ±**	r=-.117, p=.377	r=-.224, p=.090
Interaction	**r=-.464, p<.001 ±**	r=-.072, p=.588	r=-.176, p=.186
Memory	**r=-.556, p<.001 ±**	r=-.043, p=.748	r=-.172, p=.196
Control SSFT scores			
Intention	r=.033, p=.802	r=-.066, p=.619	r=-.061, p=.648
Mental State Talk	r=-.259, p=.048 *	r=-.042, p=.751	r=-.092, p=.490
Interaction	r=-.318, p=.014 *	r=.069, p=.603	r=.014, p=.917
Memory	r=-.056, p=.674	r=-.048, p=.717	r=-.059, p=.659

* p<.05, ± p<.001 Bold = remains significant after multiple comparison testing.

Table 5: Correlations (and age corrected partial correlations) between ToM SSFT scores and cognitive and empathy measures

	ToM Intention	ToM Mental State Speech	ToM Interaction
Executive Control	**r=-.438, p=.001 ±**	**r=-.389, p=.002 †**	**r=-.472, p<.001 ±**
	(r=-.189, p=.156)	(r=-.098, p=.466)	(r=-.222, p=.094)
Working Memory	**r=.495, p<.001 ±**	**r=.538, p<.001 ±**	**r=.392, p=.002 †**
	(r=.396, p=.002 †)	**(r=.446, p<.001 ±)**	(r=.266, p=.044 *)
Cognitive Empathy	r=.312, p=.016	r=.288, p=.027 *	r=.137, p=.301
	(r=.329, p.=012 *)	(r=.303, p=.021 *)	(r=.134, p=.315)
Affective Empathy	**r=.394, p=.002**	r=.154, p=.243	r=.219, p=.096
	(r=.376, p=.004 †)	(r=.106, p=.429)	(r=.178, p=.181)

* p<.05, † p<.01, ± p<.001; Age-corrected correlations in parenthesis. Bold = remains significant after multiple comparison testing.

Table 6: Results for forward stepwise regression analyses for each SSFT ToM score.

	Standardised Coefficients - Beta weights	% variance explained	ANOVA
Intention		**Total = 49.9%**	F=18.29, p<.001 [±]
Working memory	Beta=.509, p<.001	24.5%	
Affective empathy	Beta=.348, p=.001	18.5%	
Gender	Beta=-.277, p=.008	7%	
Mental State Speech		**Total = 41.7%**	F=13.13, p<.001 [±]
Working memory	Beta=.510, p<.001	28.9%	
Gender	Beta=-.264, p=.014	7.2%	
Cognitive empathy	Beta=.219, p=.039	4.5%	
Interaction		**Total = 37.9%**	F=17.11, p<.001 [±]
Executive control	Beta=-.465, p<.001	28.5%	
Education	Beta=.315, p=.005	9.4%	
Memory		**Total = 37.4%**	F=16.71, p<.001 [±]
Age	Beta=-.454, p<.001	30.9%	
Working Memory	Beta=.273, p=.020	6.4%	

[±] p<.001

Independent variables entered into models were: age, education, gender, working memory, inhibitory control, cognitive and affective empathy, and BAP traits.