British Journal of Psychology (2011)
© 2011 The British Psychological Society

The British Psychological Society

www.wileyonlinelibrary.com

The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities

Hannes Rakoczy*, Antje Harder-Kasten and Lioba Sturm
Institute of Psychology & Courant Research Centre 'Evolution of Social Behaviour'
University of Göttingen, Germany

Following up on existing, mixed findings in the literature on social cognition in old age different aspects of theory of mind were investigated in younger and older adults. In line with some previous findings, older participants – though matched with the younger ones on crystallized abilities- performed significantly worse both on tasks requiring the ascription of complex intentional attitudes to story protagonists and on tasks of recognizing subtle emotional expressions from video displays. Control analyses showed, however, that these deficits are partly explained by domain-general declines in processing speed and executive function. The implications of these findings for the nature of social cognition and its fate in old age are discussed.

How our conceptual system with which we describe and explain each other as sentient and rational actors, our *theory of mind (ToM)*, develops has been thoroughly investigated in infant and child psychology over the last three decades (see, Doherty, 2009 for review). Much research in this area documents that children acquire the ability to ascribe to others and themselves different kinds of intentional states (desires, intentions, beliefs, etc.) during early childhood (Wellman, Cross, Watson, 2001) or perhaps even in infancy (Baillargeon, Scott, & He, 2010).

How ToM fares across the lifespan is much less known, however. Only in recent years have researchers begun to investigate the development of ToM in old age, focusing in particular on two fundamental forms of ToM that are standardly distinguished: 'cognitive' ToM (involving the ascription of intentional states such as beliefs, perception, desires, etc.) and 'affective' ToM (dealing with the perception and ascription of emotional states). The few studies that exist to date have basically produced mixed and partly contradictory findings so far.

Ascribing intentional states (cognitive ToM)

The first pioneering study in this area was conducted by Happé, Winner, and Brownell (1998), using tasks in which subjects heard stories about protagonists in social

---

*Correspondence should be addressed to Hannes Rakoczy, Institute of Psychology, University of Göttingen, Waldweg 26, D-37073 Göttingen, Germany (e-mail: hannes.rakoczy@psych.uni-goettingen.de).

DOI:10.1111/j.2044-8295.2011.02040.x

interactions involving complex psychological affairs such as double bluffs, mistakes, etc. and where subjects had to make inferences about the intentional states (beliefs, etc.) of these protagonists. Their results suggested that ToM in old age may even be characterized by lifelong improvements of social understanding – akin to the growth of wisdom. Subsequent studies, however, have produced evidence either for age-related decline in social understanding or at most for lifespan stability. Maylor, Moulson, Muncer, and Taylor (2002), Charlton, Barrick, Markus, and Morris (2009), and Sullivan and Ruffman (2004), using the very same kinds of tasks as Happé *et al.* (1998) found evidence for ToM decline in elderly subjects. Duval, Piolino, Bejanin, Eustache, and Desgranges (in press) found age-related decline in a number of tasks requiring the ascription of intentional attitudes (both first- and second-order intentional states). Slessor, Phillips, and Bull (2007), in contrast, did not find any age-related differences on the level of the absolute raw values in such tasks (though they did find differences when controlling for verbal abilities; see below), nor did Keightley, Winocur, Burianova, Hongwanishkul, and Grady (2006). Similarly, McKinnon and Moscovitch (2007) did not find any age-related decline in the ability to answer simple questions about the cognitive and affective perspective of a protagonist and to ascribe first-order intentional states (but did find decline in the ascription of second-order intentional states; see below).

Ascribing emotional states (affective ToM)

A second line of research has investigated a different form of ToM, namely the ascription of emotional states, sometimes called 'affective' or 'hot' ToM. One aspect of this is the ability to perceive and interpret emotional expressions from static and/or dynamic displays. Several studies found evidence for age-related decline in the perception and interpretation of facial expressions of emotion (Bailey & Henry, 2008; Duval *et al.*, in press; Keightley *et al.*, 2006; Sullivan & Ruffman, 2004) or in detecting subtle differences in gaze direction (Slessor, Phillips, & Bull, 2008) from photos and videos.

ToM in old age: Specific decline or just general cognitive deficits?

From a theoretical point of view, the central question of interest is the following: in case there are such age-related ToM deficits, what is their basis and how are they to be explained? One possibility is that the decline – if there is such a thing – is specific to ToM. And one way to explain this, for example, would be by recourse to a prominent family of theories of the evolution and development of ToM: theories that posit a domain-specific, modular capacity solely responsible for ToM reasoning, encapsulated and functionally isolated from other cognitive faculties (e.g., Baron-Cohen & Ring, 1994; Leslie, 1994, 2005). Cognitive aging, according to such a picture, could then be seen as going along with specific degenerative processes in such a structure (candidates include the medial prefrontal cortex [mPFC] and the tempo-parietal junction [TPJ], among others; e.g., Frith & Frith, 2006).

The opposite possibility is that decline in ToM tasks merely reflects domain-general aging effects. In line with standard two-component models of lifespan cognitive development (Baltes, 1987; Horn & Cattell, 1967), for example, ToM deficits might be accounted for by general age-related decrease in cognitive abilities. One particularly relevant aspect of such general abilities might here be executive functions (EFs). From developmental and neuropsychological work, we know about very robust connections between ToM and EF: ontogenetically, ToM and EF correlate in young children highly both synchronically and diachronically (such that EF predicts later ToM) and even if controlling for extraneous factors such as chronological or mental age (e.g.; Carlson & Moses, 2001; Carlson *et al.*, 2002; Hughes, 1998; Rakoczy, 2010; Sabbagh, Moses, & Shiverick, 2006, 2006b; for an overview see Moses *et al.*, 2005). And recent neuropsychological work has shown that acquired EF deficits due to frontal lobe lesion go along with decreased performance in standard ToM tasks (Samson, Apperly, Kathirgamanathan, & Humphreys, 2005).

Which aspects of EF might matter?

Under the broad heading of EF, usually different aspects of higher cognitive functioning are subsumed. Both *a priori* task analyses (e.g., Carlson & Moses, 2001) and *a posteriori* factor analyses (Lehto, Juujaervi, Kooistra, & Pulkkinen, 2003; Miyake, Freidman, Emerson, Witzki, & Howerter, 2000) suggest that *inhibition*, *set shifting, and working memory* constitute three partly independent aspects of executive functioning. Research on the development of ToM in early childhood has found very consistent evidence that the aspects of EF most predictive of ToM are inhibition and set shifting, measured, for example, by Stroop-like inhibition tasks and Winsconsin-style card-sorting tasks, respectively (Carlson & Moses, 2001). Regarding inhibition, in the child development literature another more fine grained distinction is between so-called 'conflict inhibition' tasks and 'delay inhibition' tasks (e.g., Carlson & Moses, 2001). In pure 'delay inhibition', inhibition merely consist in suppressing one given behaviour at a given time without much further working memory requirements (delay of gratification is a typical example). In 'conflict inhibition' tasks, in contrast, a rule has to be held in mind specifying two or more potential courses of action one of which has to be inhibited according to the rule, yielding a combination of working memory and inhibition demands (examples are 'Simon says'-type tasks and Stroop-like tasks). A very consistent finding across many developmental studies with children is that regarding inhibitory tasks, 'conflict inhibition' tasks are the best predictors of ToM (e.g., Carlson, 2005; Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Sabbagh *et al.*, 2006).

For which aspects of ToM might EF matter?

Regarding the connection between ToM and EF, it has been a much disputed question in child development ToM research to which aspects of ToM EF might be related. One proposal is that EF is specifically required only in the ascription of intentional attitudes that involve some 'pull of the real', some aiming at truth. Beliefs are the paradigmatic such attitudes. Such attitudes are thought to pose inhibitory demands, in particular in ascribing false beliefs, because the ascriber has to deviate from, that is, inhibit, the normative standard (truth) pertaining to such attitudes (Sabbagh, Moses *et al.*, 2006; Henning, Spinath, & Aschersleben, 2011). Alternatively, however, it might be that EF is involved more broadly in the ascription of all kinds of intentional and emotional states to others whenever the others' perspective and one's own have to be potentially coordinated (e.g., Rakoczy, 2010).

Findings on the specificity of ToM decline so far

The evidence that speaks to the question of the specificity of ToM decline in old age so far is, like the evidence regarding the question whether there is such a decline in the first place, still rather mixed and contradictory. With regard to tasks tapping the ascription of intentional states, Maylor *et al.* (2002) found that the declining performance in the elderly in Happé *et al.*'s (1998) story tasks could not be accounted for by fluid control measures (including processing speed and fluency). Sullivan and Ruffman (2004), in contrast, found that the age effect on such story tasks disappeared once a fluid control variable (the AH4, Heim, 1970) was taken into account. And the pattern of results was slightly more complex in Slessor *et al.* (2007): though not finding any age-related differences regarding the absolute proficiency on such stories tasks, they did find age-related decline when controlling for vocabulary. However, once vocabulary was controlled for, there was an analogous (marginally significant) decline also in non-ToM control story tasks (stories without the need to draw inferences about the intentional states of the protagonists). And Duval *et al.* (in press) found that decline in the performance on simpler ToM tasks (requiring the ascription of first-order intentional states), but not on more complex tasks (involving the ascription of second-order intentional states), was mediated by declines in EF.

Three studies so far tested for age-related differences on ToM tasks requiring the ascription of intentional attitudes and systematically varied the task characteristics in terms of fluid and executive tasks demands. First, German and Hehman (2006) found that the differences between young and old adults' performance on ToM tasks (old adults were less proficient and slower) increased with the complexity of the tasks, and that age differences could be accounted for by fluid and executive control measures. Second, McKinnon and Moscovitch (2007) found that ToM performance declined with age only on more complex ToM tasks (requiring the ascription of second-order intentional states), particularly under conditions of high executive demands (in dual tasks) and that this age-related decline was not specific to ToM but extended to other kinds of social cognitive tasks (such as deontic reasoning). Third, Bailey and Henry (2008) showed that there were no age differences on ToM tasks with reduced inhibitory demands while the older subjects performed worse on tasks with higher inhibition requirements.

Regarding the perception and ascription of emotional states, the findings, again, are mixed. Sullivan and Ruffman (2004) reported that the decline in their study in recognizing emotions (in contrast to the declining performance they found in the stories tasks) could not be accounted for by fluid control variables. Similarly, Duval *et al.* (in press) reported age-related declines in emotion recognition that could not be accounted for by EF or processing speed. Bailey and Henry (2008), in contrast, found that the age-effect on emotion recognition could be accounted for to a considerable degree (though not completely) by EF measures. And Slessor *et al.* (2007) found analogous age-related deficits in emotion recognition tasks (where subjects had to read emotions from pictures and videos) as in control tasks where non-psychological factors (e.g., the age or gender of the protagonist) had to be read from pictures or videos – suggesting more general deficits, for example, in perceptual fluidity or accuracy rather than in emotion recognition.

In sum, existing findings are mixed as to the question whether and which ToM abilities are subject to age-related decline. And where such decline has been documented, findings are mixed regarding the specificity of such deficits. The present study, therefore, follows up on previous work by systematically investigating different aspects of ToM while controlling for different potentially relevant general cognitive abilities that might mediate potential age effects in ToM abilities.

The general cognitive abilities controlled for were processing speed and EF. The former was operationalized, as is standardly done, by a trail-making test (TMT). Regarding EF, we followed previous work in child development and tested those aspects of EF specifically that have previously been shown to be the best predictors of ToM: conflict inhibition (measured by two Stroop-like tasks) and set shifting (measured by a trail-making task).

The different aspects of ToM focused on here are the ascription of intentional attitudes on the one hand (as measured by Happè *et al.*'s (1998) stories tasks), and emotion recognition on the other hand (as measured by Sullivan and Ruffman's [2004] video tasks). These different aspects of ToM might be affected by age-related declines in general cognitive functioning in different ways: much developmental research with children suggests that the ability to infer and ascribe intentional attitudes, in particular beliefs (as required in the Happé *et al.* stories task) might be specifically affected by ageing effects in EF, inhibition in particular.

The ability to recognize and ascribe emotions from video scenes (as required in the Sullivan and Ruffman videos task), in contrast, might involve different components ranging from more perceptual ones (in processing the emotional expressions, body postures, etc. from the dynamic video scenes) to more cognitive ones (in inferring and ascribing the emotions of the protagonists some of which require the ascription of other mental states forming the background to complex emotions). The more perceptual components might be affected by declines in fundamental cognitive-perceptual capacity, above all processing speed, whereas the more cognitive elements (inferring mental states) might be affected by executive functioning specifically.

Method

Participants

Two groups of participants were recruited: 27 young adults (19–28 years, $M = 22.67$ years, $SD = 2.80$; 14 female) and 20 older adults (60–91 years, $M = 73.30$, $SD = 9.55$; 9 female). Younger adults were mainly psychology undergraduate students completing the study in exchange for course credit. The older adults were recruited mainly through private contacts of the authors and received small gifts for participating. All were native German speakers, had good command of their native language and had normal mental status, as indexed by scores of 28 or more in the Mini-Mental-Status-Test (Kessler, Markowitsch, & Denzler, 2000). There were no differences between the groups in their crystallized intelligence (see below).

Materials and procedure

Each participant was tested in a single session with two kinds of ToM tasks, tests of processing speed and and crystallized intelligence, and three EF tasks in the following order[1].

General cognitive measures

Processing speed. The measure for processing speed used here was the ZVT (Oswald & Roth, 1987), a German trail making test.

Crystallized intelligence. The HAWIE-R (Tewes, 1991) vocabulary subscale (a German task very similar to the Mill Hill Vocabulary Scale) was used to measure crystallized intelligence. In this task, usually considered to be a good measure of crystallized

---

[1] *The present work was part of a bigger project on psychological aging. In the context of this project, in the same session, the participants were also tested with a number of tasks/questionnaires not reported here: a questionnaire on loneliness, and a questionnaire on trait emotional intelligence.*

intelligence and an objective measure of education, participants have to explain the meaning of words (in ascending order of difficulty).

ToM tasks

ToM stories (Happé et al., 1998). German translations of the ToM stories by Happé *et al.* (1998) were used as the measure of the first aspect of ToM, namely ascribing intentional states. Participants read eight ToM stories in which protagonists were involved in complex social interactions and were subsequently asked questions regarding the intentions of the protagonist that required complex inferences about the character's mental states. In eight control stories, participants also read about complex social interactions but were then asked questions that did not require any inferences about the mentality of the story characters (but, e.g., about causal relations in the story). For each story, the participant's answer was given 0 points (incorrect), 1 point (partly correct), or 2 points (correct).

Videos task (Sullivan & Ruffman, 2004). The second aspect of ToM under study here was the recognition of emotional expressions, and the Video tasks of Sullivan and Ruffman (2004) were used as the relevant measure. Participants saw 24 short silent video clips of 2–7 s length (the original videos as used by Sullivan and Ruffman) and were instructed to choose a description (one out of two alternative words) that best fit the emotion of the person in the video. The videos were taken from various television programs and showed short sequences of everyday events. For example, in one clip, a protagonist sat on a sofa, flicking through a magazine. From the protagonist's body posture and facial expression, the participants had to infer whether she felt 'sad' or 'bored'.

EF measures

Trail-making test. The TMT (Reitan, 1958; Reitan & Wolfson, 1995) has two parts: first, in TMT-A participants have to make a trail on paper, connecting numbers in their ordinal sequence. This is a measure of pure speed, very similar to the ZVT. Second, in the TMT-B participants have to coordinate two tasks, connecting numbers and letters in alternating order (1-a-2-b, etc.). The crucial measure of EF here is a difference measure: $\text{time}_{\text{TMT-B}} - \text{time}_{\text{TMT-A}}$ (where the latter serves as a baseline measure of pure processing speed).

Stroop task. A German paper version of the Stroop task was administered (Bäumler, 1985): first, participants were asked to read aloud 30 colour words (printed in neutral black ink) (Stroop a). Second, they were asked to name the colours of 30 colour displays (Stroop b). Third came the actual test trials in which participants had to name the ink colour of 30 mismatching colour words (Stroop c). The main inhibition measure standardly used is a difference measure: $\text{time}_{\text{Stroop c}} - \text{time}_{\text{Stroop a}}$ (where the latter can be seen as a baseline measure of pure processing speed).

Day–night task (Gerstadt , Hong, & Diamond, 1994). This task consisted of 14 cards, half of them with a picture of a sun, the other with a picture of the moon and stars. The cards were shuffled, then one was flipped over at a time, and the participant's task was to say the opposite ('day' after moon pictures and 'night' after sun pictures), as fast as

Table 1. Mean scores on measures of cognitive functioning, ToM, and EF in younger and older adults

Measure	Young adults ($n = 27$) Mean (*SD*)	Older adults ($n = 20$) Mean (*SD*)	*t*	*p*
Cognitive measures				
■ Crystallized: HAWIE-R vocabulary (0–32)	24.10 (4.91)	22.89 (4.89)	.84	n.s.
■ Processing speed: ZVT time (s)	59.45 (10.34)	113.97 (37.19)	7.27	**
ToM measures				
■ ToM stories (0–16)	13.85 (1.46)	12.4 (2.16)	2.75	**
■ Control stories (0–16)	13.0 (1.83)	13.0 (1.90)	.94	n.s.
■ Video task (0–24)	19.22 (1.83)	16.70 (1.89)	4.61	**
EF measures				
(1) TMT				
■ Time $_{\text{TMT-A}}$ (s)	14.85 (2.82)	33.80 (15.05)	6.42	**
■ Time $_{\text{TMT-B}}$ (s)	38.33 (12.61)	96.40 (50.71)	5.74	**
■ Time $_{\text{TMT-B}}$ – time $_{\text{TMT-A}}$ (s)	23.48 (12.20)	62.60 (40.67)	4.74	**
(2) Stroop task				
■ Time $_{\text{Stroop A}}$ (s)	28.07 (3.36)	32.67 (4.58)	3.97	**
■ Time $_{\text{Stroop B}}$ (s)	45.48 (7.11)	49.03 (5.48)	1.86	*
■ Time $_{\text{Stroop C}}$ (s)	70.17 (16.57)	91.05 (18.55)	4.06	**
■ Time $_{\text{Stroop C}}$ – time $_{\text{Stroop A}}$ (s)	42.48 (15.11)	58.38 (15.99)	3.60	**
(3) Day–night task				
Time (s)	14.96 (2.98)	18.00 (2.64)	3.63	**
Errors	0.78 (0.75)	0.65 (0.81)	.56	n.s.

Note. *< .05; **< .01 (one-tailed).

possible. Total time in seconds to complete all 14 trials and the number of naming errors were recorded for each participant.

Results

ToM, EF, and basic cognitive abilities

Mean performances on the various cognitive tasks for the two groups are depicted in Table 1. As can be seen from the table, the younger group outperformed the older ones on a variety of tasks measuring processing speed and EFs. Most importantly, for present purposes, regarding ToM abilities, the younger group outperformed the older group on the videos task and performed significantly better than the older group on the ToM stories task but equally well regarding the control stories tasks. A separate mixed models 2 × 2 analysis of variance with age group as between-subjects factor and story type as within-subjects factor confirmed this results by revealing no main effects of age or story type, but only a significant age × story type interaction, $F(1, 45) = 4.61, p < .05, \eta_p^2 = .09$.

Correlations of cognitive, ToM, and EF measures

The correlations of the different tasks are depicted in Table 2. First of all, the two tasks measuring pure processing speed (the ZVT and the TMT-A) were highly inter-correlated ($r = .93$) and a standardized composite score was therefore formed. Similarly, following standard procedures in developmental work on EF (e.g., Carlson & Moses, 2001;

Table 2. Correlations between age, cognitive functioning, ToM, and EF measures

	HAWIE-R	Speed composite score[a]	ToM stories	Control stories	Video tasks	Inhibition I: Stroop[b]	Inhibition II: day–night (time in s)	Inhibition composite score[c]	Switching[d]	EF total composite score[e]
Age	.09	.82**	−.32*	−.03	−.52**	.53**	.41**	.59**	.64**	.70**
HAWIE-R		−.18	.18	.47**	.09	−.18	.23	.03	−.15	−.04
Speed composite score[a]			−.17	−.09	−.54**	.62**	. 28*	.56**	.87**	.78**
ToM stories				.43**	.27*	−.15	−.32*	−.29*	−.28*	−.33*
Control stories					−.01	−.05	−.07	−.08	−.19	−.14
Video tasks						−.48**	−.28*	−.48**	−.52**	−.57**
Stroop[b]							.27*	.80**	.58**	.82**
Day–night (time in s)								.80**	.21	.66**
Inhibition composite score									.50**	.92**
Time $_{\text{TMT-B}}$ – time $_{\text{TMT-A}}$ (s)										.80**

Notes. *< .05; **< .01 (one-tailed).

[a] Composite measure of the two speed tasks ZVT and TMT-A.

[b] Interference measures: time Stroop C – time Stroop A (s).

[c] Composite inhibition score of the two conflict inhibition tasks Stroop and day–night.

[d] Switching score: time $_{\text{TMT-B}}$ – time $_{\text{TMT-A}}$ (s).

[e] Composite score of all EF tasks (Stroop, day-night, and time $_{\text{TMT-B}}$ – time $_{\text{TMT-A}}$ (s)).

Table 3. Raw/partial correlations (controlling for the speed composite measure and the EF composite measure) of age and the ToM tasks

	ToM stories	Control stories	Video tasks
Age	−.32*/−.32*	−.03/.05	−.52**/−.13

Notes. *< .05; **< .01

Sabbagh *et al.*, 2006) and recent work on ToM and aging (e.g., Duval *et al.*, in press, German & Hehman, 2006), given that the measures of EF were largely inter-correlated, they were aggregated in two steps: first, at the more specific level of sub-forms of EF (conflict inhibition and switching), the two tasks tapping conflict inhibition (Stroop and day–night) were aggregated to build an inhibition composite score (the TMT-B/TMT-A score served as the unique score for switching). Second, all three tasks were aggregated to form a total EF composite score.

As can be seen from Table 2, the two ToM measures ToM stories and the Video task were somewhat correlated with each other, the ToM stories task was moderately correlated with the EF measures, and the Videos task was considerably correlated both with the speed and the EF measures.

ToM and age: Control analyses

ToM, both the ascription of intentional attitudes as measured by the ToM stories task, and the recognition of emotional expression measured in the Videos task, thus are age-related in that (1) the younger outperform the older age group and (2) age is negatively correlated with performance on these tasks. However, raw correlations and differences of means, of course, are not very informative. Three kinds of control analyses were therefore performed:

First, partial correlations of age with the ToM measure, controlling for speed and EF, were computed. As can be seen from Table 3, the correlation of age with the Video task became non-significant with these controls, whereas the correlation age-ToM stories remained significant.

Second, the two age groups were compared with analyses of covariance in which speed and EF were entered as covariates. The 2 (story type) × 2 (age group) analysis of covariance with the speed and EF composite measures as covariates retained the significant story type × age group interaction, $F(1, 43) = 4.86, p < .05, \eta_p^2 = .13$. In an analysis of covariance comparing the age groups on the Video task with the same covariates, in contrast, there did not remain a significant age effect, $F(1, 43) = 2.96, p = .09$.

Third, regression and mediator analyses (following the logic of Baron & Kenny, 1986) were run to test for the mediating influences of the extraneous cognitive factors processing speed and EF on the relation of age and ToM. As can be seen from Figure 1, age predicts both the ascription of complex attitudes in the ToM stories task and the recognition of emotions in the Videos tasks, but these effects are mediated by EF in the case of the ToM stories task, and by EF and speed in the case of the Video task[2].

---

[2] *Statistically, using the Sobel test (Preacher & Hayes, 2004), the effect proved significant in the EF – Video task mediation ($z' = 2.19$, $p < .05$). In the other two cases where including the mediator led to a decreased effect of age on the outcome variable, the effects of speed on the video task ($z' = 1.27$, $p = .21$) and of EF on the ToM stories ($z' = 1.04$, $p = .29$), the mediations failed to reach statistical significance. It should be noted, however, that the Sobel test is very conservative (e.g., MacKinnon, Warsi, & Dwyer, 1995) and given the relatively small sample size and therefore power of the present study, mediation effects are very hard to detect with this test.*

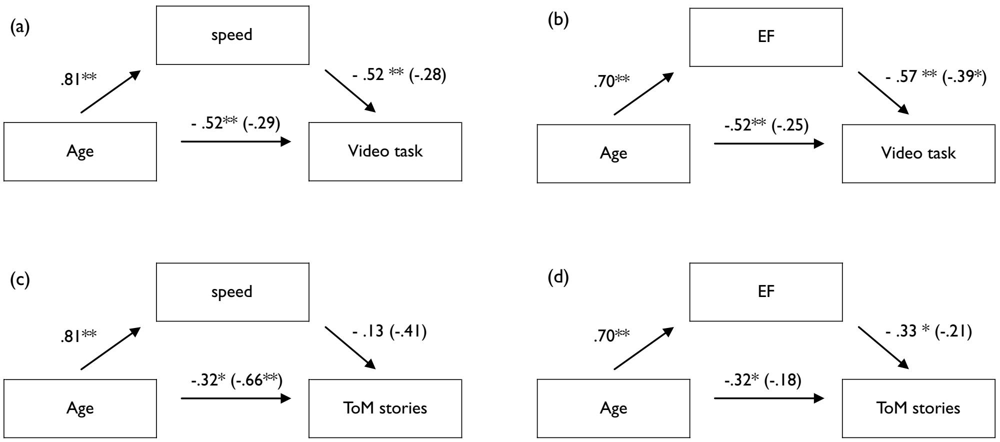

Figure 1. The mediational role of speed and EF on the different aspects of ToM (ToM stories and the Video task). Numbers represent the standardized regression coefficients without taking other predictors into account; the numbers in parentheses represent the standardized regression coefficient when age and the mediator are both used as predictors. $^{*}< .05$; $^{**}< .01$.

Discussion

Following up on previous mixed findings, the present study investigated the fate of ToM in old age. In particular, it looked at two aspects of ToM, inferring and ascribing complex intentional attitudes (measured in the ToM stories task) and recognizing and ascribing emotional states (measured with the Video task). Regarding absolute performance levels, the results were very clear: older adults, though matched with younger ones on crystallized abilities, and though performing as well as the younger ones on control stories tasks that did not require ToM, were significantly less proficient on both kinds of ToM tasks than the younger adults. These results fit with previous ones, for example, by Bailey and Henry (2008), Keightley *et al.* (2006), Maylor *et al.* (2002), Sullivan and Ruffman (2004).

Subsequent control analyses, taking into account domain-general cognitive abilities as potential mediator variables, revealed somewhat different results for the two kinds of tasks: the decline in the ability to recognize emotional expressions in the Video task quite clearly seemed to be mediated and explained by developmental changes in extraneous domain-general cognitive abilities: once general processing speed and EFs were taken into account, the group differences, the age-related correlations, and the corresponding age effects in regression analyses all disappeared. These results are inconsistent with the findings by Sullivan and Ruffman (2004) who found age-related declines in the Video task that could not be accounted for by general cognitive factors. The present results show a similar picture, in contrast, as the findings by Slessor *et al.* (2007) suggesting that the same age-related differences in the Videos task could be accounted for by more general cognitive factors (as indicated in the finding that the decline in the Video task was not specific to emotion ascription but extended to non-ToM control tasks alike).

A slightly more complicated picture emerged for the ascription of complex intentional attitudes in the ToM stories task: on the one hand, both group differences and age-related correlations remained significant even when processing speed and EF were taken into account as control variables. On the other hand, however, mediator analyses suggested that the effect of age on the ToM stories task was at least somewhat mediated by EF.

This result partly fits with previous findings by Sullivan and Ruffman (2004) that age-related differences in the ToM stories task could be accounted for by domain-general abilities (fluid abilities, however, rather than EF in their case), and fits very closely with the findings from Charlton *et al*. (2009) that decline in the ToM stories task was fully mediated by executive functioning.

The present study thus adds to a growing body of complex, partly mutually inconsistent findings regarding the development of ToM in old age. Many open and pressing questions remain thus for future work in this area. In particular, *first* of all, we need more systematic data as to which aspects of ToM reasoning are affected and to which degree by aging. So far, following the pioneering work of Happé *et al*. (1998) and Sullivan and Ruffman (2004), most studies, the present one included, have focused on a relatively limited selection of ToM tasks measuring mentalistic inferences about the intentional attitudes of story protagonists on the one hand, and subtle perceptually based emotion ascriptions on the other hand. More comprehensive, systematic and theoretically motivated batteries of ToM tasks, comparable to the ones developed for the use with children (e.g., Wellman & Liu, 2004) thus need to be devised for future research. *Second*, it remains to be seen more systematically which aspects of ToM decline in old age can be accounted for to which degree by which domain-general cognitive processes and abilities. Most studies in this area so far, again the present one included, have used relatively limited selections of EF and fluid abilities tasks as control variables. More comprehensive and theoretically motivated tests of the role of different aspects of EF, for example, like in the literature on EF and its relation to ToM in early child development (e.g., Carlson & Moses, 2001) would be needed for future work. And in addition to correlational designs like the one used in many previous studies and the present one, more direct experimental manipulations of the tasks along theoretically relevant dimensions (e.g., general processing demands; executive complexity, etc.) would be desirable. Some studies employing such methodologies have already produced first evidence that within a given task, ToM competence declines more when domain-general requirements are higher (Bailey & Henry, 2008; German & Hehman, 2006; McKinnon & Moscovitch, 2007).

One pressing theoretical challenge, finally, is the following: even if the decline in a domain-general cognitive ability such as EF perfectly explains declining performance in some ToM tasks, that still leaves open at least two fundamentally different theoretical possibilities (see, e.g., Carlson & Moses, 2001 for a discussion of these two theoretical possibilities regarding early child development): on the one hand, modularity theories could claim such a pattern reflects a pure performance problem. According to such theories, there is supposedly a ToM module that remains perfectly intact but extraneous EF abilities decline and lead to performance problems such that elderly people fail to exhibit their ToM-competence in certain types of tasks (see, e.g., Leslie, 1994, 2005 for an analogous argument regarding early development). On the other hand, competence accounts could attribute to EF a more substantial role in ToM and accordingly in its decline: rather than seeing ToM as modular and encapsulated, with EF being a mere add-on, such accounts view EF as an essential functional ingredient in ToM reasoning (see, e.g., Russell, 1996).

The existing data so far do not allow us to distinguish between these two theoretical possibilities. Future studies, therefore, should attempt to come up with viable designs to test between these two kinds of accounts more directly. One potentially relevant line of inquiry might there be found in tasks on differing levels of explicitness. Recent work with infants, for example, has found that young infants in their second year, long before they solve the standard explicit false belief task using a direct measure (Wimmer & Perner, 1983), indicate some implicit sensitivity to others' beliefs in studies with indirect measures using violation of expectation habituation methods (e.g., Onishi & Baillargeon, 2005) and spontaneous anticipatory looking (e.g., Southgate, Senju, & Csibra, 2007). Modularity theorists interpret such findings as revelation of early competence masked by performance problems arising in explicit tasks (e.g., Leslie, 2005). Interestingly, a recent study found that normally developing adults show such spontaneous (probably rather involuntary) anticipatory looking behaviour, but that adults with Asperger Syndrome do not – although they solve explicit false belief tasks (Senju, Southgate, White, & Frit, 2009). One modularity theoretical interpretation of such a finding is that the hypothesized ToM module of people with Asperger Syndrome functions differently in rather basic ways, and that the competence they acquire in explicit ToM tasks is based on different processed and acquired via different routes than in normally developing adults. Against the background of such findings, it would be interesting and potentially theoretically relevant to see how ToM fares in old age on tasks varying in explicitness, in particular whether ToM deficits extend to such spontaneous indirect tasks.

Another line of inquiry that might be relevant pertains to ToM in relation to other forms of social reasoning. Prominent modularity theories posit a bundle of independent modules dealing with different aspects of social cognition, ToM on the one hand, but separate from this at least another module for reasoning about social exchanges and the detection of cheaters (e.g., Cosmides & Tooby, 1992). Were this true, should age-related developments in old age in such different aspects of social cognition be fundamentally separate and independent from each other. More systematic data on the common or independent fate of different aspects of social reasoning might therefore be illuminating. One existing study has already moved in such a direction and suggests that different forms of social reasoning are affected similarly by cognitive aging and argues on that basis against strict modularity accounts (McKinnon & Moscovitch, 2007). It remains a somewhat open question, however, as the authors note, whether this pattern of results reveals merely a common influence of domain-general factors on the *performance* in different types of social cognition tasks, or whether it documents truly common cognitive bases in the *competence* regarding different forms of social cognition.

Finally, the most challenging puzzle to be resolved in this area concerns the bigger picture regarding the relation of losses and gains in different form of social cognition. While now the majority of studies coming from a ToM perspective, including the present one, suggest some decline in some aspects of social cognition in old age, research from a rather different and separate tradition emphasizes lifelong positive developments in social reasoning related to wisdom (e.g., Sternberg & Jordan, 2005). In a recent study, for example, Grossmann *et al.* (2010) found that in wisdom-related tasks requiring reasoning about social conflicts and their resolution, the older people outperformed the younger ones in the use of higher order reasoning schemas, in emphasizing multiple perspectives of different actors, and in dealing with the limits of certainty and knowledge – abilities that look suspiciously like practical forms of ToM reasoning. So how are these different lines of research with their different focus and the different findings to be reconciled? The challenge for future work here lies in delineating the difference between forms of social reasoning that are sensitive to cognitive aging and those that are not – to spell out, in other words, a kind of fluid-crystallized distinction for social cognition.

Acknowledgements

We would like to thank Marlen Kaufmann for helping in organizing the study, Ted Ruffman and Ellen Winner for sharing their material and their friendly cooperation. This work was supported by a 'Dilthey Fellowship' of the Volkswagen Foundation and the Fritz Thyssen Foundation and by the German Initiative of Excellence.

References

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*, *63B*(4), 219–P226.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

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

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Baron-Cohen, S., & Ring, H. (1994). A model of the mindreading system: Neuropsychological and neurobiological perspectives. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 183–207). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Bäumler, G. (1985). *Farbe-Wort-Interferenztest (FWIT) nach JRStroop*. Toronto, Zürich, Göttingen: Hogrefe.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73–92.

Charlton, R. A., Barrick, T. R., Markus, H. S., & Morris, R. G. (2009). Theory of mind associations with other cognitive functions and brain imaging in normal aging. *Psychology and Aging*, *24*(2), 338–348.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Doherty, M. J. (2009). *Theory of mind: How children understand others' thoughts and feelings*. New York, NY, US: Psychology Press.

Duval, C., Piolino, P., Bejanin, A., Eustache, F., & Desgranges, B. (in press). Age effects on different components of theory of mind. *Consciousness and Cognition*.

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*(3), 189–198.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*, *101*(1), 129–152.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old an a stroop-like day-night test. *Cognition*, *53*, 129–153.

Grossmann, I., Na, J., Varnuma, M. E. W., Park, D. C., Kitayama, S., & Nisbett, R. E. (2010). Reasoning about social conflicts improves into old age. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7246–7250.

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

Heim, A. W. (1970). *The AH4 group test of intelligence*. Windsor, UK: NFER-Nelson.

Henning, A., Spinath, F. M., & Aschersleben, G. (2011). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology*, *108*(3), 513–531.

Horn, J. L., & Cattell, R. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*(2), 107–129.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*(2), 233–253.

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*(3), 558–572.

Kessler, J., Markowitsch, H. J., & Denzler, P. (2000). *Mini-Mental-Status-Test (MMST)*. Göttingen: Beltz Test GMBH.

Lehto, J. H., Juujaervi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59–80.

Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*(1–3), 211–238.

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*(10), 459–462.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41–62.

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

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.

Miyake, A., Freidman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The untiy and diversity of executive functions and their contribution to complex "frontal lobe" tasks: A latentn variable analysis. *Cognitive Psychology*, *41*, 49–100.

Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2005). On the specificity of the relation between executive function and children's theories of mind. In W. Schneider, R. Schuhmann-Hengsteler & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 131–145). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Onishi, K. H., & Baillargeon, R. e. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Oswald, W. D., & Roth, E. (1987). *Zahlen-Verbindungs-Test (ZVT)* (2nd ed.). Göttingen: Hogrefe.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731.

Rakoczy, H. (2010). Executive function and the development of belief-desire psychology. *Developmental Science*, *13*(4), 648–661.

Reitan, R. M. (1958). Validity of the trail making test as an indication of organic brain damage. *Perceptual and Motor Skills*, *8*(8), 271–276.

Reitan, R. M., & Wolfson, D. (1995). Category test and trail making test as measures of frontal lobe functions. *Clinical Neuropsychologist*, *9*, 50–60.

Russell, J. (1996). *Agency: Its role in mental development*. Hove: Erlbaum (UK) Taylor & Francis Ltd.

Sabbagh, M., Moses, L., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, *77*(4), 1034–1049.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The Development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*(1), 74–81.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain: A Journal of Neurology*, *128*(5), 1102–1111.

Senju, A., Southgate, V., White, S., & Frit, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–885.

Slessor, G., Phillips, L. H., & Bull, R. (2007). Exploring the specificity of age-related differences in theory of mind tasks. *Psychology and Aging*, *22*(3), 639–643.

Slessor, G., Phillips, L. H., & Bull, R. (2008). Age-related declines in basic social perception: Evidence from tasks assessing eye-gaze processing. *Psychology and Aging*, *23*(4), 812–822.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592.

Sternberg, R. J., & Jordan, J. (2005). *A handbook of wisdom: Psychological perspectives*. New York, NY, US: Cambridge University Press.

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

Tewes, U. (1991). *HAWIE-R. Hamburg-Wechsler Intelligenz Test für Erwachsene- Revision*. Bern: Huber.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Received 5 January 2011; revised version received 11 March 2011