Nice to Meet You – Adult Age Differences in Empathic Accuracy for Strangers

Elisabeth S. Blanke, Antje Rauers, and Michaela Riediger

Max Planck Institute for Human Development, Berlin, Germany

Author Note

Elisabeth S. Blanke, Antje Rauers, and Michaela Riediger, Max Planck Research Group "Affect Across the Lifespan," Max Planck Institute for Human Development, Berlin, Germany.

This research was funded by the Max Planck Society. The data reported in this manuscript were collected for Elisabeth S. Blanke's doctoral dissertation. We would like to thank the MaxNetAging Research School for providing Elisabeth S. Blanke with financial and academic support. We would like to express our gratitude to Dulce Erdt, Kristin Bischof, Caroline Cohrdes, Tobias Gfesser, Sandy Jahn, Nora Koster, Thomas Lennefer, Claudia Medel, Vanessa Petruo, Anika Schulz, Konrad Senf, and Linda Wuttke for their help with

data collection, and Amy Michéle for editorial assistance. We would also like to thank our participants who brought this study to life.

Correspondence concerning this article should be addressed to Elisabeth S. Blanke, Max Planck Institute for Human Development, Max Planck Research Group “Affect Across the Lifespan,” Lentzeallee 94, 14195 Berlin, Germany. Email: blanke @mpib-berlin.mpg.de.

Abstract

Empathic accuracy is the ability to correctly identify others' thoughts and feelings. Based on evidence from past laboratory experiments, researchers concluded that this ability decreases throughout adulthood. This conclusion, however, was mostly based on evidence regarding isolated components of the ability to read others' thoughts and feelings (e.g., inferring thoughts or feelings from facial expressions presented without context). In contrast, empathic accuracy involves the integration of a multitude of such inferences from diverse sources of information that are available in everyday interactions (e.g., facial and bodily expressions, prosody, communication content, situational context, etc.). To strengthen empirical evidence on age differences in this integrative ability, we assessed empathic accuracy in dyadic interactions between 102 younger (20-31 years) and 106 older (69-80 years) women, paired in same-age or mixed-age dyads. In these interactions, older women were only less empathically accurate than younger women when judging their interaction partner's negative feelings and when judging thoughts that accompanied experiences of negative affect. In contrast, there were no age differences in empathic accuracy for positive feelings and for thoughts accompanying experiences of positive affect. These results were independent of the age of the interaction partner. The current study thus provides further evidence that age differences in empathic accuracy (a) may be qualified by situational properties, such as valence of inferred content, and (b) can be less pronounced when integration of multiple sources of information is possible than research investigating isolated information channels has thus far suggested. *(241/ 250)*

KEYWORDS: empathic accuracy; age differences; dyadic interaction *(3/5)*

Words: 8190 (+296 words footnotes)

Nice to meet you – Adult age differences in empathic accuracy for strangers

Imagine that someone you do not recognize is looking at you in the subway. You may wonder, “Is there something on my face?”, “Does this person know me?”, or simply, “What is this person thinking?” Your answers to these questions likely guide your reaction (e.g., examining your reflection in the window, looking away, or smiling). People frequently encounter such situations in which they try to infer what other people, familiar or unfamiliar, think and feel. The degree to which these inferences are correct has been referred to as empathic accuracy (Ickes, Stinson, Bissonnette, & Garcia, 1990). Previous research suggested that such empathic accuracy is beneficial for social adjustment, for example, as reflected in higher marital satisfaction or a better ability to provide social support (e.g., Cohen, Schulz, Weiss, & Waldinger, 2012; Verhofstadt, Buysse, Ickes, Davis, & Devoldre, 2008).

Given the importance that abilities involved in making inferences about others’ thoughts and feelings have for social interactions, prior empirical evidence that they might decline with age is unsettling (for meta-analyses see Henry, Phillips, Ruffman, & Bailey, 2013; Ruffman, Henry, Livingstone, & Phillips, 2008). Most of this evidence stems from age-comparative research on specific components of the ability to correctly identify others’ thoughts and feelings studied in isolation from each other, such as the ability to read emotional facial expressions when presented without context. Empathic accuracy, however, is more than just the combination of such isolated skills. As Ickes (1997, p.2) put it, empathic accuracy is achieved by “complex psychological inference in which observation, memory, knowledge, and reasoning are combined to yield insight into the thoughts and feelings of others.” It thus reflects the empathizer’s ability to integrate the manifold and complex pieces of information that usually are available in natural interactions and that stem from different sources (e.g., facial and bodily expressions, prosody,

communication content, situational context, etc.). To date, little is known about older adults' empathic accuracy as it derives from such complex integration of information. In the following, we first review evidence from age-comparative research on isolated skills that are conceptually related to empathic accuracy, and then turn to the relatively scarce findings on adult age differences in empathic accuracy, as both strands of research informed our hypotheses.

Age differences in isolated skills that are related to empathic accuracy

Empirical research on adult age differences in skills that are related to empathic accuracy points to a decrease in the abilities to read emotional expressions and to understand mental states. Results from a meta-analysis (Ruffman et al., 2008) showed that older adults perform worse than younger adults in reading the majority of emotional expressions displayed in faces, voices, and bodily postures – an ability that is referred to as emotion recognition (see also Mill, Allik, Realo, & Valk, 2009; Ruffman, Murray, Halberstadt, & Taumoepeau, 2010). In emotion-recognition tasks, participants are usually presented with posed emotional expressions from one isolated information channel (e.g., facial expressions). The participant's task typically is to select the expression intended by the poser (the target) from a number of response options. Another recent meta-analysis by Henry et al. (2013) showed an age-related decrease in the ability to infer mental states, referred to as Theory of Mind (ToM). Like emotion-recognition tasks, ToM tasks often consist of multiple-choice questions, asking participants to choose a mental state that best describes a (usually posed) expression in a picture or video or a mindset of a fictional character in a story. The causes for age-related decreases in emotion recognition and ToM are not yet understood. Age-related structural (e.g., cognitive and neurophysiological) and motivational changes as well as differences in the facial cues (e.g., from the mouth versus eye region) used to judge emotional expressions have been discussed as potential candidates (Charles & Campos,

2011; Isaacowitz & Stanley, 2011; Ruffman, 2011; Ruffman et al., 2008). Empirical support for these lines of reasoning is, however, still rare (Ruffman, 2011).

Moreover, the ways of measuring these skills, especially emotion recognition, have been criticized as being potentially disadvantageous for older adults. The two most prominent criticisms have been the lack of age-fairness and the lack of ecological validity (e.g., Isaacowitz & Stanley, 2011; Rauers, Blanke, & Riediger, 2013; Richter, Dietzel, & Kunzmann, 2011; Richter & Kunzmann, 2011; Riediger, Studtmann, Westphal, Rauers, & Weber, 2014). Age-fairness is lacking because stimuli usually included younger and middle-aged adults' emotional expressions as targets, but not older adults' (but see, e.g., Riediger, Voelkle, Ebner, & Lindenberger, 2011 for an exception). This may have disadvantaged older perceivers in that it has been hypothesized that older adults might perform better at recognizing emotional expressions of their own age group, typically referred to as own-age advantage, although empirical evidence is not yet conclusive (e.g., Ebner, He, & Johnson, 2011; Riediger et al., 2011). The ecological validity in most previous emotion-recognition tasks has been limited for several reasons: The stimulus material often consisted of still pictures displaying posed emotional expressions without any contextual information. Spontaneous emotional expressions, in contrast, are dynamic and rapidly changing; and they are more subtle than posed expressions (Reisenzein, Studtmann, & Horstmann, 2013). Artificiality of the task might especially hinder older adults' performance, as older adults seem to profit in other domains from familiar and less artificial tasks (e.g., cognitive tasks; Kliegel, Martin, McDaniel, & Phillips, 2007). Furthermore, older adults might benefit more from context information in emotion recognition than younger adults do (Noh & Isaacowitz, 2013). Ruffman (2011) argued that even when more realistic video tasks were used, the majority of empirical studies nevertheless pointed to a decline in emotion

recognition with advancing adult age. There are a few exceptions, however, and older adults have occasionally even been found to outperform younger adults when judging affective experiences from video material on affective expressions (e.g., Sze, Goodkind, Gyurak, & Levenson, 2012). Criticism on the lack of ecological validity has also been expressed for ToM tasks that have typically used pictures or videos of posed mental states or faux pas, or sometimes comic strips or vignettes (Dziobek, 2012). Accordingly, older adults might be disadvantaged in ToM tasks as well. In their meta-analysis, Henry et al. (2013), however, reported that older adults performed worse than younger adults regardless of the task being used, including dynamic, audio-visual ToM video tasks.

To summarize, methodological shortcomings might contribute to age-related differences in emotion recognition and ToM, but they are not likely to fully explain them. Moreover, the ability to read thoughts and feelings in daily life is likely to be more than the sum of the isolated skills that have been measured in emotion recognition and ToM tasks. There certainly are situations in daily life in which it is important to read feelings from isolated sensory channels (e.g., when having a conversation with a stranger on the phone). Most everyday social interactions, however, require integrating multiple pieces of information from diverse channels such as facial and bodily expressions, prosody, and the content of the conversation, to name a few examples. Unlike the majority of the previous studies, we were therefore interested in empathic inferences that require this complex integration of skills, that is in empathic accuracy and in potential age-related differences therein.

Age differences in empathic accuracy

Empathic accuracy is defined as the correct inference of others' thoughts and feelings and operationalized as the concordance between the self-report of a target person who experiences

thoughts and feelings and the respective judgment of the empathizer (e.g., Ickes et al., 1990). It is usually measured in live interactions or using already videotaped situations (Rollings, Cuperman, & Ickes, 2011). Both paradigms address the methodological criticism raised on emotion recognition and ToM tasks. Ecological validity is enhanced in these tasks as they target naturally occurring thoughts and feelings within a situational context. Age-fairness can be achieved by varying the age of the partners in a dyadic interaction or the target persons in the videotapes. To the best of our knowledge, the focus of studies looking at age differences in empathic accuracy until now was limited to the ability to infer feelings, not thoughts. The assessment of empathic accuracy for thoughts requires the analysis and coding of open-answer formats. It is therefore usually more difficult to implement (and in some settings not feasible) than the assessment of empathic accuracy for feelings, which can be measured with rating scales (Ripoll et al., 2013). Ickes (2011) nevertheless made a compelling case for the importance of investigating empathic accuracy for thoughts as well, showing that participants spontaneously reported more thoughts than feelings (Ickes & Cheng, 2011). This suggests that thoughts represent an important facet of people's inner experiences. Furthermore, empathic accuracy for thoughts and empathic accuracy for feelings sometimes diverge: In a study aimed at improving empathic accuracy of graduate students by using feedback, only empathic accuracy for feelings, not for thoughts, was enhanced after several weeks of training (Barone et al., 2005).

Few studies specifically investigated adult age differences in empathic accuracy for feelings. Results suggest that empathic accuracy for feelings, like emotion recognition and ToM, declines with age, and that specific features of the tasks may moderate these age differences. In a video-task study, age differences only emerged when older adults judged the feelings of a target person who talked about a topic that was presumably of little motivational significance for older

adults, but not when the topic was age-relevant, suggesting that motivational factors contribute to age differences in empathic accuracy (Richter & Kunzmann, 2011). In another study, Richter et al. (2011) assessed empathic accuracy for feelings using context-poor (without sound) or context-rich (audio-visual) videos depicting happy, sad, and angry targets. In this study, younger adults outperformed older adults in the inference of sadness and anger in both conditions, but not in the context-rich condition of happiness. The authors argued that older adults might have been more motivated to accurately judge the positive content than the negative, which is in line with the so called "positivity effect" in aging research. The positivity effect describes a tendency of older adults to be less sensitive to negative information and/or more sensitive to positive information, presumably to regulate (maintain or enhance) their emotional well-being (e.g., Carstensen & Mikels, 2005).

Both studies used video tasks to assess empathic accuracy. Although video tasks maximize internal validity with standardized assessment, they cut back on ecological validity. To investigate empathic accuracy in real-life contexts, Rauers et al. (2013) used experience sampling and showed that younger romantic couples inferred their partner's feelings in daily life more accurately than older couples did – but only when the partner was present at that time, not when the partner was absent. The authors argued that empathic accuracy in the absence of the partner was informed by knowledge about the partner, whereas in the presence of the partner, it was also informed by sensory cues such as facial emotional expressions. This study also emphasizes that age differences in empathic accuracy occur not only when judging the feelings of unfamiliar persons, but also when judging close emotional partners. A limitation of the study was the restricted age range of the partner's age, as partners in each couple were of similar ages. None of the studies included empathic accuracy for thoughts, thus missing a potentially

important facet of interpersonal inferences (Ickes, 2011). In the current study, we therefore built on previous findings concerning age differences in empathic accuracy for feelings and related skills. It also should be noted that whenever empathic accuracy for feelings is assessed separately (not together with thoughts), it is usually measured using emotion rating scales or rating dials instead of the open-answer format that was used in the classic studies conducted by Ickes. We addressed the methodological limitations of previous research discussed here and extended our research scope to incorporate empathic accuracy for thoughts.

The present study: Hypotheses

The aim of the current study was to strengthen empirical evidence on age differences in empathic accuracy for thoughts and feelings. To enhance ecological validity within the controlled environment of a laboratory, we used a dyadic interaction task to investigate interactions between unfamiliar persons that talked about personal experiences. We assumed that talking about personal events that the participants could pick themselves would make this an emotionally relevant and motivating task for older as well as for younger adults. To enhance age-fairness and generalizability, we systematically varied the age-group composition of the dyads and had younger and older adults interact either with members of the same, or members of the other age group. Another novel addition to the research on age differences in empathic accuracy was the parallel assessment of accuracy for thoughts and feelings as recommended by Ickes (2011). To our knowledge, this is the first empirical attempt to test age differences in the ability to read others' thoughts and feelings in a live interaction between unfamiliar younger and older adults. Based on empirical evidence on age differences in empathic accuracy for feelings, as well as in emotion recognition and ToM, we expected younger adults to be more accurate in judging the thoughts (Hypothesis 1) and the feelings (Hypothesis 2) of an unfamiliar interaction partner.

Furthermore, we exploratively followed up on previous research (Richter et al., 2011) that found that older adults' empathic-accuracy level was equal to younger adults' level when a context-rich and ecologically valid task was provided, but only when the stimulus material was positively valenced, not when it was negatively valenced. This was interpreted as a motivational process with which older adults try to maintain their emotional well-being. We assumed that the maintenance of well-being should also be a goal in an interaction with an unfamiliar person and were therefore interested in whether this pattern would also be found in a more realistic interaction task.

Method

Participants

Participants were 208 women from the Berlin area recruited from the participant pool of the Max Planck Institute for Human Development, Berlin, Germany, as well as from an online advertisement in the internet. We tested two age groups: n = 102 younger adults (age range = 20–31 years, M = 25.95, SD = 3.06), and n = 106 older adults (age range = 69–80 years, M = 72.94, SD = 2.52).[1] Participants were fluent in German and the sample was approximately stratified according to education, with 59% of the younger and 50% of the older adults holding a German university entrance qualification (Abitur). Participants were recruited independently from one another and were asked at first encounter whether they knew their assigned interaction partner, which was never the case. One half of the sample was paired with participants from the other age group (52 mixed-age dyads) and the other half, with partners from the same age group (52 age-homogeneous dyads: 25 younger and 27 older dyads). We chose to investigate same-sex dyads (all participants were female) to reduce the complexity of the research design, which was optimized for the purpose of investigating age differences. Participants were told in advance that

there would be a videotaped conversation and they gave their informed consent. The study consisted of two sessions, but only the first session is relevant for the current research. Participants received 50 Euro as compensation for both sessions. The study was approved by the ethics committee of the Max Planck Institute for Human Development.

Procedure

After having given written consent and answering a short questionnaire, participants had five minutes to introduce themselves to each other, and to get accustomed to the camera that was already recording them. They were then asked to think of one recent personal event during which they had felt particularly bad (e.g., angry, tense, sad, or unhappy) and another event during which they had felt particularly good (e.g., excited, happy, content, or balanced). They were told that their conversation would be about these events, with the aim of getting to know each other. Participants had a few minutes to think of events and were given the opportunity to put down notes. During the following conversation, which lasted twelve minutes, the partners took turns in talking about their events, with three minutes for each event at their disposal. All partners were instructed to listen and, if they liked, to ask questions and make comments. Signals for the time were given by an audiotaped instruction because the experimenter left the room for the duration of the recordings. Participants started with the negative events and ended with the positive ones because we did not want the participants to have negative feelings at the end of the task.[2] Within the mixed-age subsample, we ensured that younger and older women started the conversations equally often; otherwise who started the conversation was randomly assigned. After filling out further questionnaires that were not relevant for the current research question and a short break, participants watched the video of their interaction, with interruptions at eight time points (“tape stops”). These tape stops were pre-defined by the experimenter, based on a time-contingent

criterion: In each three-minute segment of the twelve-minute conversation, a tape stops was set at approximately one minute and at approximately two minutes, resulting in eight stops in total. The experimenters had a tolerance margin of about 15 seconds before and after the predefined mark. He or she defined the tape-stops at the end of naturally occurring interaction segments (e.g., the end of a sentence). This was done to prevent unnatural interruptions in the participants' speech (and potential distortions of the meaning of a sentence). At each tape stop, participants used the items described in the next section (self-report) to indicate their most important thought as well as their feelings at that particular point in the interaction. The participants then watched the video a second time, this time reporting what the conversation partner might have thought and felt at the same tape stops (judgment).

This procedure of assessing empathic accuracy was similar to the "unstructured dyadic interaction paradigm" introduced by Ickes and colleagues (e.g., Rollings et al., 2011). However, we modified the procedure in several important ways: (a) the participants knew that they were being filmed (vs. being unaware of the recording), (b) the topic and timing of the conversation was semi-structured (vs. unstructured), (c) the tape stops were pre-defined (vs. freely chosen by the participants), and (d) participants were instructed to report their own thoughts and the assumed thoughts of their partner using an open-answer format; and to separately report their own feelings and the assumed feelings of their partner using rating scales (vs. reporting both thoughts and feelings together in an open-answer format). The modifications in terms of filming and structure of the conversation were made to enhance compliance and enable an emotional exchange between the strangers. We decided to pre-define the tape stops to keep the number of tape stops constant across participants. We predefined identical tape stops for both partners of a dyad to be able to statistically consider that the participants may project their own emotions onto

the other person, which we will explain in more detail in the Methods section. To separate emotional states from thoughts and to make the study comparable to other age-differential studies on empathic accuracy, we specifically asked the participants to report thoughts using an open-answer format and to rate their feelings using emotion rating scales. In the following, we will refer to measures derived from the open-answer responses as "empathic accuracy for thoughts," and to measures derived from the emotion ratings as "empathic accuracy for feelings."

Measures

Self-reported thoughts and judgments of thoughts. While watching the recording of their interaction, participants wrote down their own thoughts at each tape stop, using an open answer format. Participants were asked to focus on the most important thought that they were having in the conversation at that particular moment when the tape stop had occurred. When watching the recording a second time, participants reported the most important thought they assumed their partner had had at that time, again using an open answer format.

Self-reported feelings and judgment of feelings. Participants rated their own feelings and their judgments of the partner's feelings for each of the eight tape stops using nine affect items (see below). A 7-point scale ranging from 0 *(not at all)* to *6 (very much)* was used. The items were selected to cover facets of low- as well as high-arousal positive and negative affect.

Emotional valence of thoughts and feelings. Affective experiences are often categorized according to their valence into positive and negative experiences (e.g., Russell, 1980; Watson & Tellegen, 1985). This distinction has also been used for the present data and analyses. Positive affect was represented by five items (PA: happy, excited, content, comfortable, and balanced), and negative affect by four items (NA: nervous, sad, uncomfortable, and tense). The

positive and negative affect items were averaged, yielding separate scores for self-reported positive and negative affect at each tape stop (average PA of personal means: younger: M = 3.30, SD = 1.00; older: M = 3.21, SD = 0.93; average NA of personal means: younger: M = 1.38, SD = 0.99; older: M = 1.40, SD = 0.79). At each tape stop, reliability for the self-reported affect measures was good, ranging from α = .86 to α = .92 for positive and α = .70 to α = .82 for negative affect. The same aggregation was used for the judgment of the partner's feelings (average PA of personal means: younger: M = 3.29, SD = 0.89; older: M = 3.25, SD = 0.95; average NA of personal means: younger: M = 1.15, SD = 0.76; older: M = 1.24, SD = 0.71). The reliability for the judgment of partner's positive feelings was good (ranging from α = .87 to α = .91); for the judgment of partner's negative feelings reliability was acceptable (ranging from α = .68 to α = .79).

To determine the emotional experience that accompanied the self-reported thoughts, we used the personal mean of each participant for her self-reported positive and negative affect across the eight tape stops. Thought entries were regarded as accompanying positive affect when the self-reported positive affect score for that entry was higher than or equal to the personal mean of this person (positive thoughts). Following the same logic, thought entries were regarded as accompanying negative affect when the negative affect score was higher than or equal to the personal mean of this person (negative thoughts).[3] On average, 51% of each younger woman's thoughts (SD = 11) and 54% of each older woman's thoughts (SD = 12) were categorized as accompanying positive affect; and 41% (SD = 13) of the younger women's thoughts and 42% (SD = 13) of the older women's thoughts were categorized as accompanying negative affect.

Empathic accuracy for thoughts. Two trained coders who were blind to the hypotheses rated the similarity between the empathizer's judgment and the partner's self-reported thoughts

at each tape stop. A 3-point coding system (Ickes et al., 1990) was used: essentially different content (coded 0); somewhat similar, but not the same, content (coded 1); essentially the same content (coded 2). Throughout the coding process, extensive consensus meetings were regularly held with the first author, in which disagreement between the coders was discussed until consensus was reached. Inter-rater reliability was computed using coders' ratings before the consensus meetings. Because Category 2 (essentially the same content) was only coded 14 times out of 1648, Categories 1 and 2 were collapsed into one category that we interpret as indicating that the empathizer correctly inferred the thoughts of her interaction partner at that tape stop. The inter-rater-reliability was $\kappa = .69$, which may be considered a good reliability, given the complexity of the coding (Wirtz & Caspar, 2002).

For each empathizer, consensus coding was used to obtain an average proportion of the correctly inferred thoughts the partner had had while feeling more positive than usual (positive thoughts) or negative than usual (negative thoughts). The resulting empathic accuracy measures for thoughts had a theoretical range between 0 (none of the interaction partner's thoughts were correctly inferred) and 1 (all of the interaction partner's thoughts were correctly inferred). These measures were severely positively skewed. We therefore repeated our analyses with inverse transformed accuracy scores to approach normality. This transformation did not change the pattern of results. For reasons of parsimony and interpretability, we thus analyzed and report the untransformed scores (mean accuracy for positive thoughts: younger: $M = 0.25$, $SD = 0.21$; older: $M = 0.24$, $SD = 0.22$; mean accuracy for negative thoughts: younger: $M = 0.21$, $SD = 0.22$; older: $M = 0.15$, $SD = 0.21$). Three participants (two older women and one younger woman) were excluded from the analyses of empathic accuracy for thoughts because they were unwilling or unable to report any thoughts. One older woman had a very high score in empathic accuracy

for negative thoughts (as compared to the other older women). Adjusting this score did not change the results; we therefore used the unadjusted data.

Empathic accuracy for feelings. For feelings, we chose two approaches to model empathic accuracy that we describe in more detail below: a person-level approach (comparable to empathic accuracy for thoughts) and an additional situation-level approach that allowed us to model empathic accuracy at each tape stop. We used this additional approach to demonstrate that both methodological approaches (person-level and situation-level) yield the same results for the analysis of age differences in empathic accuracy for feelings. We applied the model only to feelings (not to thoughts) because it required interval-scaled data at each measurement occasion and that was not available for the thought data.

Person-level approach. Empathic accuracy for feelings was calculated as the Pearson correlation between the judgment of the empathizer and the self-report of the partner over the eight tape stops, yielding one score per person (person-level). As correlations generally do not follow a normal distribution, we used Fisher's z-transformation to approach normality and to make the correlations comparable across participants (Kenny, 1987). This z-transformed measure of empathic accuracy for feelings therefore had a theoretical rage of approximately +/− 3 (equal to $r \approx +/- 1$), with higher positive scores representing higher accuracy (positive feelings: younger: $M = 0.90$, $SD = 0.57$; older: $M = 0.82$, $SD = 0.60$; negative feelings: younger: $M = 0.86$, $SD = 0.53$; older: $M = 0.69$, $SD = 0.62$).

In the analysis of empathic accuracy for negative feelings, four participants (three older women and one younger woman) were excluded because they did not display any variation in their judgment of partner's negative affect ratings. One younger woman had a very low score in empathic accuracy for positive feelings (as compared to the other younger women). Similarly,

one younger and one older woman had very low scores in empathic accuracy for negative feelings (as compared to their own age group). Adjusting these scores did not change the results; we therefore used the unadjusted data.

Situation-level approach. In this approach, we modeled empathic accuracy at each tape stop using the truth and bias model of judgment (West & Kenny, 2011). Empathic accuracy for feelings was represented as the prediction of each empathizer's judgment by the partner's self-reported feelings across the tape stops, controlling for the empathizer's own current feelings. Higher estimates for the partner's self-reported feelings represent higher empathic accuracy. Negative affect ratings displayed positive skewedness, but because an inverse transformation that we applied with the aim to approach normality did not change the pattern of results, we again analyzed untransformed and unadjusted scores for reasons of parsimony. One observation (i.e., one tape stop for one empathizer) was missing because one older participant only completed seven of the eight judgments of partners' affect.

Methodological considerations concerning dyadic dependency

The evaluation of thoughts and feelings of one partner was likely to be influenced by the other partner, as can be expected in dyadic interactions (Kenny, Kashy, & Cook, 2006). We therefore analyzed the data using the actor-partner interdependence model (APIM; Kenny et al., 2006). The APIM controls for the dependency of the dyad members as it allows for correlated dependent variables within the dyads. The situation-level approach (concerning empathic accuracy for feelings) additionally required taking within-person interdependencies into consideration, which we will describe later with the results of these analyses. The dyad members were treated as being indistinguishable because no meaningful characteristic such as gender or social roles could be used to order the two persons in the dyad (Kenny et al., 2006). In our

sample, all participants were female, all of them were empathizers as well as partners, and only in half of the sample did partners differ by age group (for a similar analytical approach, see West, Dovidio, & Pearson, 2014). All analyses were carried out using the MIXED procedure of the SAS/ STAT ® software, Version 9.2 of the SAS System for Windows (SAS Institute, 2011). We followed the recommendations by Kenny et al. (2006) and estimated all models using restricted maximum likelihood (REML) and the Satterthwaite approximation of degrees of freedom.

Results

We hypothesized that older women would be less empathically accurate than younger women in inferring their interaction partners' thoughts (Hypothesis 1) and feelings (Hypothesis 2). We first report results for the person-level analyses of empathic accuracy for thoughts and for feelings, followed by an additional situation-level analysis of empathic accuracy for feelings.

Age differences in empathic accuracy for thoughts

We used the actor-partner interdependence model (APIM), implemented in a multilevel model with dyads members as repeated measurements within the dyads. We predicted the person-level empathic accuracy score of the empathizer for thoughts by the age group of the empathizer. Age group of the empathizer was coded as -1 for younger adults and 1 for older adults. We differentiated between thoughts accompanied by above-average positive or negative affect, respectively. Consistent with our hypothesis, the age group of the empathizer was a significant predictor of empathic accuracy for thoughts accompanying negative affect (estimate = −0.033, SE = 0.015, p = .023, intercept = 0.180, SE = 0.017, p < .001). Relative to the average score of younger women, older women's scores were therefore about one third lower. There were no age differences for the inference of thoughts accompanying positive affect (estimate =

−0.011, *SE* = 0.015, *p* = .445; intercept = 0.246, *SE* = 0.016, *p* < .001). In other words, younger women outperformed older women in empathic accuracy only for thoughts accompanying above-average negative, but not positive affect (see Figure 1). To test whether the partner's age-group membership or the age composition of the dyad was also related to empathic accuracy, we further controlled for the age group of the partner and the interaction between both partners' age groups (age group of partner was coded as *−1* for younger and *1* for older women). When controlling for the partner's age-group membership as well as the interaction between the empathizer's and the partner's age groups, the effect of the empathizer's age stayed robust for thoughts accompanying negative affect (estimate = −0.031, *SE* = 0.015, *p* = .041) and remained non-significant for thoughts accompanying positive affect (estimate = −0.007, *SE* = 0.015, *p* = .649). For both types of thoughts, there was no significant effect of partner's age group (positive: estimate = 0.028, *SE* = 0.015, *p* = .066; negative: estimate = 0.010, *SE* = 0.015, *p* = .505) or of the interaction between empathizer's age group and partner's age group (positive: estimate = 0.010, *SE* = 0.016, *p* = .526; negative: estimate = −0.008, *SE* = 0.017, *p* = .642). In other words, only the age-group membership of the empathizer predicted empathic accuracy for thoughts, but neither the age-group membership of the partner nor the age composition of the dyad.

Age differences in empathic accuracy for feelings (person-level approach)

We again used the APIM to predict the empathizer's person-level empathic accuracy score for feelings with the age group of the empathizer. Again, age group of the empathizer was coded as *-1* for younger adults and *1* for older adults. We differentiated between positive and negative affect. Age group of the empathizer was a significant predictor of empathic accuracy for negative feelings (estimate = −0.087, *SE* = 0.040, *p* = .033; intercept = 0.775, *SE* = 0.042, *p* < .001). Relative to the average score of younger women, older women's scores were therefore

about one fifth lower. There were no age differences in the ability to identify the partner's positive feelings (estimate = −0.041, *SE* = 0.038, *p* = .285; intercept = 0.859, *SE* = 0.047, *p* < .001). Corresponding to the results for empathic accuracy for thoughts, this indicates that younger women outperformed their older counterparts only in the inference of negative feelings, not positive ones. Figure 2 shows empathic accuracy scores for positive and negative feelings separately for younger and older participants. When we again controlled for the age group of the partner and the interaction between both partners' age groups, the effect of the empathizer's age stayed robust for negative feelings (estimate = −0.088, *SE* = 0.040, *p* = .031) and remained non-significant for positive feelings (estimate = −0.036, *SE* = 0.041, *p* = .380). For both positive and negative feelings, there was no significant effect of partner's age group (positive: estimate = 0.015, *SE* = 0.041, *p* = .716; negative: estimate = −0.031, *SE* = 0.040, *p* =.446) or the interaction between the empathizer's age group and the partner's age group (positive: = −0.025, *SE* = 0.047, *p* = .597; negative: estimate = −0.066, *SE* =0.042, *p* =.118). This indicates that, again, only the age group membership of the empathizer predicted empathic accuracy for feelings, but not the age-group membership of the partner or the age composition of the dyad.

Age differences in empathic accuracy for feelings (situation-level approach)

In this analysis, we used a modified APIM, this time treating the eight tape stops as repeated measures nested within the dyads. We accounted for interdependencies arising from the repeated measurements over the eight tape stops and the dyadic interdependencies among the partners' ratings at a given tape stop by implementing a correlated residual structure using the Kronecker product structure[4] (Bolger & Shrout, 2007; Kenny et al., 2006). Which partner started the conversation was used as a random grouping factor, dividing the dyads into two equally sized groups (Olsen & Kenny, 2006). To assess age differences in empathic accuracy for feelings on a

situation-level, we used the truth and bias model from West and Kenny (2011). In this model, the judgment of the empathizer is predicted by the partner's self-report (the "truth") at each of the eight tape stops, while controlling for the self-report of the empathizer (the "bias"), because assumed similarity between own affect and partner's affect might affect accuracy (Kenny & Acitelli, 2001). Following West and Kenny (2011), we centered the empathizer's and the partner's self-reports (predictors) and the empathizer's judgment (dependent variable) at the personal mean of the partner's self-rating (subtracting the personal mean of the partner's self-ratings from each individual rating). We further entered age group of the empathizer as a main effect (coded *−1* for younger and *1* for older adults) and—to test our hypothesis—the interaction between the self-report of the partner ("truth") and the age group of the empathizer. Again, we performed separate analyses for positive and negative affect. Parameter estimates of these models testing age differences in empathic accuracy for positive and negative affect on the situation level are shown in Table 1.

The situation-level analysis revealed that feelings of the empathizer were a significant predictor for the empathizer's judgment of the partner's feelings, indicating that participants' own emotional experiences contributed to their judgments: When they, for example, rated themselves as one unit more positively (positive affect ranging from 0–6), they rated their partner as 0.394 units more positively at the same time (all other predictors being equal). Above and beyond this assumed similarity, the partner's self-rated feelings predicted the empathizer's judgment, indicating that empathic accuracy of the empathizers was significantly different from zero. In other words, when the partner, for example, rated herself as one unit more positively, the empathizer rated her 0.315 units more positively as well, indicating that the partner's feelings were systematically related to the empathizer's judgment. The two-way interaction between the

partner's self-rated feelings and the empathizer's age group was significant for negative affect, but not for positive affect. In line with the person-level analysis, the estimates indicate higher empathic accuracy of younger than of older women for negative, but not for positive feelings. The intercept for the judgment of negative affect was significantly lower than zero. This indicates that, overall, participants underestimated their partners' negative affect (see West & Kenny, 2011, p.364, for details). The intercept for the judgment of positive affect did not significantly differ from zero, indicating that participants did not systematically over- or underestimate their partners' positive affect.

There was no main effect of the empathizer's age group on the judgment. This means that there was no evidence for differences between younger and older empathizers' judgments of their partners' positive or negative affect; neither of the age groups provided higher or lower judgments than the other group. We then additionally controlled for partners' age-group membership (two-way interaction between the partner's self-rated affect and her age group) and age-group composition of the dyad (three-way interaction between partner's self-rated affect, empathizer's age group, and partner's age group) as well as all lower level effects. The two-way interaction between partner's self-rated affect and empathizer's age group stayed significant for negative feelings (estimate = −0.063, *SE* = 0.022, *p* = .005) and remained non-significant for positive feelings (estimate = −0.017, *SE* = 0.020, *p* = .395). The two-way interaction between partner's self-rated affect and partner's age was not significant (positive: −0.007, *SE* = 0.020, *p* = .721; negative: −0.016, *SE* = 0.022, *p* = .484), nor was the three-way interaction between partner's self-rated affect, empathizer's age, and partner's age significant (positive: −0.014, *SE* = 0.020, *p* = .495; negative: −0.024, *SE* = 0.022, *p* = .287). This again indicates that neither the

partner's age group nor the age-group composition of the dyad influenced empathizer's empathic accuracy.

Discussion

The aim of the current study was to extend empirical evidence on adult age differences in the abilities to read others' thoughts and feelings. Previous research on emotion recognition and Theory of Mind (ToM) points to a decrease of these abilities with age (Henry et al., 2013; Ruffman et al., 2008), but researchers have recently criticized the methods widely used to measure those skills as lacking in age-fairness and ecological validity (e.g., Isaacowitz & Stanley, 2011). To investigate age differences in empathic accuracy, we therefore used a dyadic interaction paradigm that captures the ability to accurately judge naturally occurring thoughts and feelings. We expected older women in our sample to perform worse than younger women in empathic accuracy for thoughts (Hypothesis 1) and for feelings (Hypothesis 2). We also explored the role of emotional valence in empathic accuracy.

Consistent with our hypotheses, we found younger women to be more accurate than their older counterparts in describing their partners' thoughts when these thoughts were accompanied by high levels of negative affect ("negative thoughts"). Younger women were also more accurate in inferring their partners' negative feelings. These findings are in line with the reviewed empirical findings on age differences in empathic accuracy as well as related constructs, such as emotion recognition and ToM. To assess empathic accuracy for feelings, we replicated and extended our results in a second analytic approach at the level of each tape stop, additionally controlling for the self-reported feelings of the empathizer as a potential bias. Above and beyond any assumed similarity between the partners, younger women were still more accurate than older women in judging their partners' negative feelings. Age group of the partner and the age-group

composition of the dyads were not related to empathic accuracy for negative thoughts or feelings, emphasizing the pivotal role of the empathizer's age. In sum, these results provide further evidence that age differences in the abilities to infer negative thoughts and feelings can be observed in paradigms with enhanced age-fairness and ecological validity. Age differences in negative affect (NA) occurred reliably with different analytic approaches (i.e., a coding procedure as well as a correlational approach) and levels of analysis (i.e., person-level as well as situation-level).

We did not find age differences in empathic accuracy for the partner's positive feelings or thoughts that where accompanied by high levels of positive affect ("positive thoughts"). These results for positive affect (PA) fit in with results from a study by Richter et al. (2011), who also did not find age differences in empathic accuracy for feelings when a context-rich video task that featured a positive event was used. The authors assumed that older adults were more motivated to process positive than negative material. This explanation alludes to the "positivity effect", which denotes a motivational shift towards increasingly preferring positive over negative information with increasing adult age (e.g., Carstensen & Mikels, 2005). This effect has been observed in attention, memory, and decision-making tasks (Scheibe & Carstensen, 2010). Although the positivity effect does not fully explain age differences in emotion-recognition tasks (Ruffman, 2011), it might still influence older adults' performance in these tasks, maybe even more so in paradigms that demand personal motivational involvement like our empathic accuracy task. We therefore consider it possible that older adults paid closer attention to their interaction partner's positive as opposed to their partner's negative affect, which may have resulted in higher empathic accuracy for positive feelings and for thoughts that were accompanied by above-average positive affect than for negative feelings and thoughts.

Concerning the valence of the thoughts, it is important to note that the valence was assigned on the basis of the accompanying emotions (rather than on the content of the thought)—it is therefore possible that in some cases, the content of the thoughts themselves was not particularly negative (e.g., when a person felt tense while disclosing a positive memory to the unfamiliar interaction partner). This may suggest that older women's interest in the accurate judgment of the conversation partner's thoughts decreased when they sensed that her mood was negative. Additionally, it is possible that already the instruction to the interaction partner to talk about something negative may have potentially triggered older empathizers to be less attentive compared to when the interaction partner was instructed to talk about a positive event. It is further important to note that the situation-level analysis for empathic accuracy for feelings did not reveal age-differential patterns of systematic over- or underestimation of positive or negative affect, meaning that older adults did not simply attribute more positive or less negative affect to their partners than younger adults did. Both younger and older women underestimated the partner's negative affect.

From the age differences in empathic accuracy for feelings one can only deduce that older women were, on average, less accurate than younger women at inferring the affective intensity of the partner's negative feelings. This difference, however, was unsystematic, that is, there was no difference between younger and older participants in their bias toward underestimating their partner's negative affect. It remains an open question whether older adults were more likely than younger adults to mistake certain negative emotions for other negative emotions. We also do not know how exactly older adults were inaccurate concerning empathic accuracy for negative thoughts. As empathic accuracy for thoughts was assessed with an open-answer format, there were more ways to be inaccurate for thoughts than there were for feelings

(which were assessed with rating scales). Empathic accuracy for thoughts may therefore have been the more difficult task. Although in line with Richter et al. (2011), our findings differ from the age-related decrease in the inference of positive as well as negative emotions that has usually been observed in emotion-recognition tasks. Although age differences in emotion recognition of positively valenced stimuli are usually smaller than those observed for negatively valenced stimuli, Ruffman (2011) considered this a methodological artifact: Emotion-recognition tasks typically feature only one or two positive emotions, but many negative ones, which might make the positive emotions easier to distinguish in a multiple choice answering format. As our participants did not judge distinct emotional expressions, this possible artifact does not explain our findings.

Participants in the present study interacted with an unfamiliar person. Being confronted with an unfamiliar and therefore rather unpredictable person might have motivated them to monitor their interaction partner more carefully to understand and get to know her. In this context, it also seems possible that older adults' preference for positive emotions might become particularly salient: Whereas older adults might be motivated to perceive negative feelings and the accompanying thoughts of close social partners (e.g., to provide support), they might not feel the same way about the negative feelings and accompanying thoughts of an unfamiliar person. In our study, older participants might have been (consciously or unconsciously) motivated to enhance the pleasantness of the short interaction, thus being more attentive towards detecting positive as opposed to negative thoughts and feelings. Our pattern of results is consistent with predictions of the positivity effect because age differences only emerged in the inference of negative thoughts and feelings, not positive ones. We did not, however, measure the motivation to attend to positive or negative feelings and their accompanying thoughts directly in the sense of

a positivity effect. We therefore cannot rule out alternative or complementary explanations for our findings. It is, for example, possible that age-related neurophysiological changes caused the decline in empathic accuracy for negative material (Ruffman, 2011; Ruffman et al., 2008), as an age-related loss of gray matter in the medial prefrontal cortex has been shown to be associated with a decline in the recognition of fear (Williams et al., 2006).

Limitations and Outlook

We followed the call for more ecologically valid and age-fair empirical research on age differences in the abilities to read others' thoughts and feelings and investigated empathic accuracy in dyadic interactions between younger and older women. Although we tried to maximize ecological validity, our paradigm differed from a realistic conversation in many important aspects (e.g., provision of a general topic, time limit, and camera recording). Another limitation of our design is our all-female sample; thus we cannot rule out that there might have been gender differences in our interaction task. Furthermore, our results pertain only to the interaction between unfamiliar persons. Familiarity of social partners has been shown to raise levels of empathic accuracy (Stinson & Ickes, 1992) and researchers have been debating that age differences in the inference of thoughts and feelings might be attenuated or even disappear when judging a familiar partner (Henry et al., 2013; Isaacowitz & Stanley, 2011). At least for empathic accuracy for feelings, Rauers et al. (2013) have shown that in the presence of their romantic partners, older adults judged their partners' feelings in daily life less accurately than younger adults did. This emphasizes the existence of age differences in empathic accuracy even in close social relationships. As pointed out earlier in the discussion, these age differences in various social relationships might be differentially motivated. Another important limitation of the study is the cross-sectional design, as age-group differences do not necessarily correspond to intra-

individual change with age. Furthermore, we chose to investigate two age groups only (20–31 years and 69–80 years). Future research is needed to compare performances in empathic accuracy over the lifespan. As it was not the aim of our study to determine the reasons for an age-related decline in empathic accuracy, the question of why age differences in empathic accuracy occur when they do occur is still an open research question. Furthermore, it is still open for investigation to what extent this relative inaccuracy of older adults is adaptive or maladaptive. Although most people usually want to know what others' think and feel, there are situations in which accuracy does not help, or even damages social relationships (e.g., Ickes & Simpson, 2007; Myers & Hodges, 2009), especially concerning negative emotional content (e.g., Elfenbein & Ambady, 2002). In terms of the stranger on the subway that we introduced in the beginning, it might not be helpful or might even be distressing to accurately understand negative thoughts and feelings of an unfamiliar person on the train. On the other hand, there are situations in which older adults' inaccuracy, especially for negative thoughts and feelings, seems to be maladaptive, for example when trying to detect deceit (e.g., Ruffman, Murray, Halberstadt, & Vater, 2012; Stanley & Blanchard-Fields, 2008). A pivotal question is therefore how these age differences in the judgment of negative thoughts and feelings relate to older adults' socio-emotional adjustment in daily life.

Conclusion

In our interactive paradigm, younger women outperformed older women only in empathic accuracy for negative, but not for positive, feelings and accompanying thoughts. These findings are consistent with a motivational perspective on age differences in empathic accuracy as they point to the positivity effect, although we cannot rule out alternative explanations for this differential effect. The current study is in line with recent studies showing that adult age

differences in empathic accuracy might not be as consistent as past research on emotion recognition and ToM has suggested. It underscores that age differences in empathic skills are not universal, but qualified by situational factors, such as emotional valence.

Footnotes

[1] The sample size was originally set to 100 younger and 100 older adults. We oversampled two age-mixed dyads and two age-homogenous older dyads (resulting in 2 additional younger and 6 additional older participants) due to various reasons which are explained throughout the manuscript (e.g., inability to report thoughts, disregarding of the timing of the task) and due to one person quitting the study after the first session (the second session was not relevant for the current research question).

[2] Two dyads (two age-homogeneous dyads, one young-young and one old-old) did not completely adhere to the timing of the task. Excluding these two dyads from the analyses did not change our results and we therefore chose to keep them as a part of the sample. Two other participants from two different dyads (one younger woman from a young-young dyad and one older woman from an old-old dyad) were not able or did not want to report a negative event. They instead talked about the reasons why there was nothing negative in their lives. Excluding these dyads from the analyses did not change our results, and we again chose to keep them as a part of the sample. Excluding all of these four dyads who did not completely adhere to the task did also not change the pattern of results.

[3] This procedure led to a partial overlap between "positive" and "negative" thoughts (average within-person occurrence of overlapping episodes: younger: M = 7%, SD = 10; older: M = 7%, SD = 12). The result pattern did not change when we repeated our analyses while excluding tape stops that had been categorized as positive and negative at the same time (e.g., tape stops at which the participant experienced mixed affect).

[4] TYPE = UN@AR(1) in SAS

Table 1

Actor-Partner Interdependence Model (Multilevel Modeling) Predicting Empathizer's Judgment of Partner's Positive and Negative Affect (N = 1663 observations)

	Positive Affect		Negative Affect	
Predictors	*estimate*	*SE*	*estimate*	*SE*
Constant	0.010	0.053	−0.193**	0.057
Empathizer's Affect	0.394**	0.020	0.222**	0.021
Partner's Affect	0.315**	0.021	0.354**	0.022
Empathizer's Age Group	0.057	0.054	−0.044	0.058
Partner's Affect x Empathizer's Age Group	−0.021	0.019	−0.066**	0.022

Note. We report all effects as unstandardized regression coefficients. *$p < .05$, **$p < .01$. The estimate for "Partner's Affect" reflects empathic accuracy. The estimate for "Partner's Affect x Empathizer's Age Group" reflects age differences in empathic accuracy.

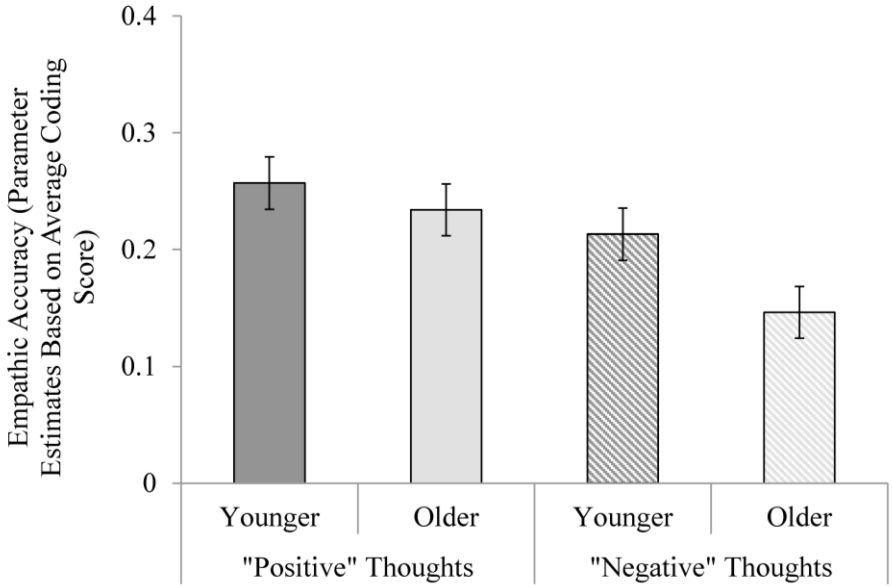

Figure 1. Empathic accuracy (model results) for partner's thoughts, subdivided according to valence. "Positive" and "negative" thoughts represent thoughts that the partner had while feeling more positive/negative than her personal average. Error bars represent ±2 *SE*. To obtain standard errors per age group, we ran complementary models with reversed dummy codes (1/0).

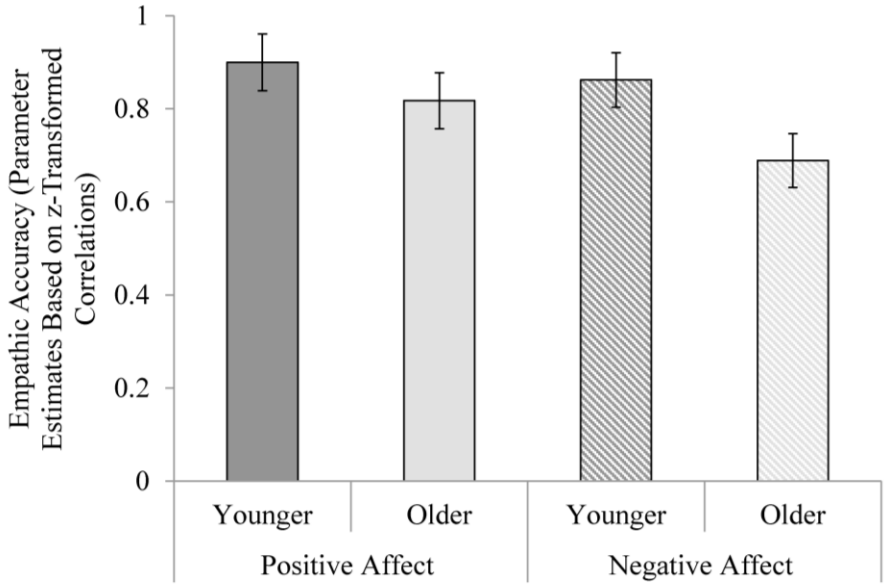

Figure 2. Empathic accuracy (model results) for partner's feelings, subdivided according to valence. Error bars represent ±2 *SE*. To obtain standard errors per age group, we ran complementary models with reversed dummy codes (1/0).

References

Barone, D. F., Hutchings, P. S., Kimmel, H. J., Traub, H. L., Cooper, J. T., & Marshall, C. M. (2005). Increasing empathic accuracy through practice and feedback in a clinical interviewing course. *Journal of Social and Clinical Psychology, 24*, 156-171. doi: 10.1521/jscp.24.2.156.62275

Bolger, N., & Shrout, P. E. (2007). Accounting for statistical dependency in longitudinal data on dyads. In T. D. Little, J. A. Bovaird & N. A. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development* (pp. 285-298). Mahwah, NJ: Erlbaum

Carstensen, L. L., & Mikels, J. A. (2005). At the intersection of emotion and cognition: Aging and the positivity effect. *Current Directions in Psychological Science, 14*, 117-121. doi: 10.1111/j.0963-7214.2005.00348.x

Charles, S. T., & Campos, B. (2011). Age-related changes in emotion recognition: How, why, and how much of a problem? *Journal of Nonverbal Behavior, 35*, 287-295. doi: 10.1007/s10919-011-0117-2

Cohen, S., Schulz, M. S., Weiss, E., & Waldinger, R. J. (2012). Eye of the beholder: The individual and dyadic contributions of empathic accuracy and perceived empathic effort to relationship satisfaction. *Journal of Family Psychology, 26*, 236-245. doi: 10.1037/a0027488

Dziobek, I. (2012). Comment: Towards a more ecologically valid assessment of empathy. *Emotion Review, 4*, 18-19. doi: 10.1177/1754073911421390

Ebner, N. C., He, Y., & Johnson, M. K. (2011). Age and emotion affect how we look at a face: Visual scan patterns differ for own-age versus other-age emotional faces. *Cognition & Emotion, 25*, 983-997. doi: 10.1080/02699931.2010.540817

Elfenbein, H. A., & Ambady, N. (2002). Predicting workplace outcomes from the ability to eavesdrop on feelings. *Journal of Applied Psychology, 87*, 963. doi: 10.1037//0021-9010.87.5.963

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. doi: 10.1037/a0030677

Ickes, W. (2011). Everyday mind reading is driven by motives and goals. *Psychological Inquiry, 22*, 200-206. doi: 10.1080/1047840X.2011.561133

Ickes, W., & Cheng, W. (2011). How do thoughts differ from feelings? Putting the differences into words. *Language and Cognitive Processes, 26*, 1-23. doi: 10.1080/01690961003603046

Ickes, W., & Simpson, J. A. (2007). Motivational Aspects of Empathic Accuracy. In G. J. O. Fletcher & M. S. Clark (Eds.), *Blackwell Handbook of Social Psychology: Interpersonal Processes* (pp. 229-249). Malden: Blackwell Publishers Ltd.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 59*, 730-742. doi: 10.1037/0022-3514.59.4.730

Isaacowitz, D. M., & Stanley, J. T. (2011). Bringing an ecological perspective to the study of aging and recognition of emotional facial expressions: Past, current, and future methods. *Journal of Nonverbal Behavior, 35*, 261-278. doi: 10.1007/s10919-011-0113-6

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. Boston: Little Brown and Company.

Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology, 80*, 439-448. doi: 10.1037/0022-3514.80.3.439

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Kliegel, M., Martin, M., McDaniel, M. A., & Phillips, L. H. (2007). Adult age differences in errand planning: The role of task familiarity and cognitive resources. *Experimental Aging Research, 33*, 145-161. doi: 10.1080/03610730601177395

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion, 9*, 619-630. doi: 10.1037/a0016562

Myers, M. W., & Hodges, S. D. (2009). Making it up and making do: Simulation, imagination, and empathic accuracy. In K. Markman, W. Klein & J. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 281-294). New York, NY: Psychology Press.

Noh, S. R., & Isaacowitz, D. M. (2013). Emotional faces in context: Age differences in recognition accuracy and scanning patterns. *Emotion, 13*, 238-249. doi: 10.1037/a0030234

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods, 11*, 127-141. doi: 10.1037/1082-989X.11.2.127

Rauers, A., Blanke, E., & Riediger, M. (2013). Everyday empathic accuracy in younger and older couples: do you need to see your partner to know his or her feelings? *Psychological Science, 24*, 2210-2217. doi: 10.1177/0956797613490747

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review, 5*, 16-23. doi: 10.1177/1754073912457228

Richter, D., Dietzel, C., & Kunzmann, U. (2011). Age differences in emotion recognition: The task matters. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 66B*, 48-55. doi: 10.1093/geronb/gbq068

Richter, D., & Kunzmann, U. (2011). Age differences in three facets of empathy: Performance-based evidence. *Psychology and Aging, 26*, 60-70. doi: 10.1037/a0021138

Riediger, M., Studtmann, M., Westphal, A., Rauers, A., & Weber, H. (2014). No smile like another: Adult age differences in identifying emotions that accompany smiles. *Frontiers in Psychology, 5*. doi: 10.3389/fpsyg.2014.00480

Riediger, M., Voelkle, M. C., Ebner, N. C., & Lindenberger, U. (2011). Beyond "happy, angry, or sad?": Age-of-poser and age-of-rater effects on multi-dimensional emotion perception. *Cognition and Emotion, 25*, 968-982. doi: 10.1080/02699931.2010.540812

Ripoll, L. H., Zaki, J., Perez-Rodriguez, M. M., Snyder, R., Strike, K. S., Boussi, A., . . . New, A. S. (2013). Empathic accuracy and cognition in schizotypal personality disorder. *Psychiatry research, 210*, 232-241. doi: 10.1016/j.psychres.2013.05.025

Rollings, K. H., Cuperman, R., & Ickes, W. (2011). Empathic accuracy and inaccuracy. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 143-156). New York, NY: Wiley & Sons.

Ruffman, T. (2011). Ecological validity and age-related change in emotion recognition. *Journal of Nonverbal Behavior, 35*, 297-304. doi: 10.1007/s10919-011-0116-3

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews, 32*, 863-881. doi: 10.1016/j.neubiorev.2008.01.001

Ruffman, T., Murray, J., Halberstadt, J., & Taumoepeau, M. (2010). Verbosity and emotion recognition in older adults. *Psychology and Aging, 25*, 492-497. doi: 10.1037/a0018247

Ruffman, T., Murray, J., Halberstadt, J., & Vater, T. (2012). Age-related differences in deception. *Psychology and Aging, 27*, 543-549. doi: 10.1037/a0023380

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178. doi: 10.1037/h0077714

SAS Institute. (2011). The SAS system for Windows. Release 9.2. SAS Inst. Cary, NC.

Scheibe, S., & Carstensen, L. L. (2010). Emotional aging: Recent findings and future trends. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *65*, 135-144. doi: 10.1093/geronb/gbp132

Stanley, J. T., & Blanchard-Fields, F. (2008). Challenges older adults face in detecting deceit: The role of emotion recognition. *Psychology and Aging, 23*, 24-32. doi: 10.1037/0882-7974.23.1.24

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, *62*, 787-797. doi: 10.1037/0022-3514.62.5.787

Sze, J. A., Goodkind, M. S., Gyurak, A., & Levenson, R. W. (2012). Aging and emotion recognition: Not just a losing matter. *Psychology and Aging, 27*, 940-950. doi: 10.1037/a0029367

Verhofstadt, L. L., Buysse, A., Ickes, W., Davis, M., & Devoldre, I. (2008). Support provision in marriage: The role of emotional similarity and empathic accuracy. *Emotion, 8*, 792-802. doi: 10.1037/a0013976

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219-235. doi: 10.1037/0033-2909.98.2.219

West, T. V., Dovidio, J. F., & Pearson, A. R. (2014). Accuracy and bias in perceptions of relationship interest for intergroup and intragroup roommates. *Social Psychological and Personality Science, 5*, 235-242. doi: 10.1177/1948550613490966

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review, 118*, 357-378. doi: 10.1037/a0022936

Williams, L. M., Brown, K. J., Palmer, D., Liddell, B. J., Kemp, A. H., Olivieri, G., . . . Gordon, E. (2006). The mellow years?: Neural basis of improving emotional stability over age. *The Journal of Neuroscience, 26*, 6422-6430. doi: 10.1523/jneurosci.0022-06.2006

Wirtz, M., & Caspar, F. (2002). *Beurteileruebereinstimmung und Beurteilerreliabilitaet. [Interrater agreement and interrater reliability.]*. Goettingen: Hogrefe.