COGNITION AND EMOTION, 2018
https://doi.org/10.1080/02699931.2018.1454403

Routledge
Taylor & Francis Group

BRIEF ARTICLE

Check for updates

Sex differences in facial emotion perception ability across the lifespan

Sally Olderbak[a], Oliver Wilhelm[a], Andrea Hildebrandt[b] and Jordi Quoidbach[c]

[a]Institute for Psychology and Education, Ulm University, Ulm, Germany; [b]Institute for Psychology, Ernst Moritz Arndt Universität Greifswald, Greifswald, Germany; [c]Department of People Management and Organisation, ESADE Business School, Barcelona, Spain

ABSTRACT
Perception of emotion in the face is a key component of human social cognition and is considered vital for many domains of life; however, little is known about how this ability differs across the lifespan for men and women. We addressed this question with a large community sample ($N = 100,257$) of persons ranging from younger than 15 to older than 60 years of age. Participants were viewers of the television show "Tout le Monde Joue", and the task was presented on television, with participants responding via their mobile devices. Applying latent variable modeling, and establishing measurement invariance between males and females and across age, we found that, for both males and females, emotion perception abilities peak between the ages of 15 and 30, with poorer performance by younger adults and declining performance after the age of 30. In addition, we show a consistent advantage by females across the lifespan, which decreases in magnitude with increasing age. This large scale study with a wide range of people and testing environments suggests these effects are largely robust. Implications are discussed.

ARTICLE HISTORY
Received 20 December 2017
Revised 13 February 2018
Accepted 14 March 2018

KEYWORDS
Emotion perception; emotion recognition; sex differences; aging; latent variable modeling

Emotion perception is considered a key component of our communicative socio-emotional skills and vital for many domains of life, facilitating engagement at work and academic performance (Mayer, Caruso, & Salovey, 2016). While research suggests that older adults are worse at perceiving emotions relative to younger adults (e.g. Isaacowitz et al., 2007; Mill, Allik, Realo, & Valk, 2009; for a meta-analytic review see also Ruffman, Henry, Livingstone, & Phillips, 2008) and women perform better than men (Thompson & Voyer, 2014), little is known about how emotion perception abilities differ across the life span separately for men and women. Moreover, studies that have looked at age-related differences have methodological limitations including underpowered samples, the use of measures with known psychometric limitations, and no evaluation of measurement invariance between groups, which is a prerequisite for unbiased estimates of age differences. The present study addresses these limitations.

Emotion perception

Emotion perception is broadly defined as the ability to discriminate and identify emotions expressed by another person, either vocally, facially, or through the body (Mayer et al., 2016), with tests of this ability most often targeting the ability to identify *facial* expressions of emotion. Cognitive testing confirmed that emotion perception, when measured specifically with the face modality, is a unique cognitive ability, distinct yet strongly related with the ability to perceive faces with neutral expressions (face perception), the ability to remember faces with neural expressions (face memory), the ability to remember faces with emotional expressions (emotion memory), and general mental ability (Hildebrandt, Sommer, Schacht, & Wilhelm, 2015). Also, persons better in facial emotion perception ability are generally better at facilitating thought using emotion, understanding emotions, managing emotions (e.g. MacCann, Joseph, Newman, & Roberts, 2014), and expressing emotions (Elfenbein & Eisenkraft, 2010).

CONTACT Sally Olderbak sally.olderbak@uni-ulm.de
Supplemental data for this article can be accessed at https://doi.org/.10.1080/02699931.2018.1454403.

Aging

The decline in fluid abilities (e.g. mental speed, working memory) across the adult lifespan is well documented (e.g. Lövdén & Lindenberger, 2005) with the age at which declines start, as well as the rates of decline, depending on the specific fluid ability (Hartshorne & Germine, 2015). Given strong relations between emotion perception and face perception, one would expect emotion perception abilities to decline at a similar rate as face perception abilities, starting around 60 years of age (Hildebrandt, Herzmann, Sommer, & Wilhelm, 2010).

Emotion perception involves activation of the amygdala, insula, striatum (Haxby & Gobbini, 2011), and the ventromedial prefrontal cortex (Mather, 2016), although the extent to which these regions are recruited differs between older and younger adults (e.g. Gunning-Dixon et al., 2003). While these structures have functional and structural decline with increasing age, this decline is less relative to other brain areas. Thus, an aging brain is not considered an exhaustive reason for poorer performance and generally, the true cause is unclear (Mather, 2016). One possible explanation is older adults use different search patterns, relative to younger adults, which may partially explain the differences in performance (e.g. Sullivan, Campbell, Hutton, & Ruffman, 2017). This may be related to the posterior-anterior shift associated with aging, a process of brain dedifferentiation showing younger adults utilise occipital brain areas (including the fusiform face area) more during face perception, relative to older adults (Park et al., 2004).

Sex

Sex differences in emotion perception ability is also well documented (Thompson & Voyer, 2014), with females routinely showing an advantage over males, even since infancy, implying this advantage cannot be due solely to the effects of socialisation (McClure, 2000). However, the extent to which sex differences are maintained across the age span is unknown. A recent meta-analysis showed the largest sex differences were for teenagers and young adults (13–30 years), with smaller sex difference for children (<13 years of age) and persons older than 30 years (Thompson & Voyer, 2014). However, a large scale study of over 7000 adults from 18 to 75 years old found no interaction of age and sex in emotion perception (Sasson et al., 2010). Those results are comparable to findings regarding sex differences in the perception of neutral faces, where female advantage remains stable from young to older adults, and sex differences in memory for faces, where female advantage also remains stable (18–88 years; Sommer, Hildebrandt, Kunina-Habenicht, Schacht, & Wilhelm, 2013).

Current study

An understanding of sex-specific trends in emotion perception performance is important for thinking about sex-specific patterns in how the brain ages. Likewise, since emotion perception is considered one of the basic abilities underlying more complex socio-emotional abilities (Mayer et al., 2016), for which there is little to no information about how they behave with increased age, these results will provide an initial framework.

We examine emotion perception ability for males and females across the lifespan in a large community sample. This study goes beyond others in several ways. First, this study has a sufficiently powered sample to detect the expected small effect sizes (Ruffman et al., 2008; Thompson & Voyer, 2014). Second, we apply latent variable modeling to provide a more precise estimate of performance and to establish measurement invariance, in order to ensure the comparability of mean performance estimates between groups. Finally, we present detailed comparisons of mean performance for several age groups, providing a more detailed understanding of this ability across the lifespan for males and females.

Methodology

Sample

Participants were viewers of the television show "Tout le Monde Joue" [translated to Everybody Plays] hosted on the public French television channel France 2, which aired February 28, 2017. Of the 2.5 million viewers (7% of France's total population, 11.6% of people who have a television), 158,127 viewers participated in the television-presented tasks via their mobile devices (smartphone and tablets). Of this sample, 100,257 persons provided data relevant to this study. There were somewhat more females (63%) than males and the sample ranged in age from below 15 years (6%), 15–30 years (26%), 31–45 years (31%), 46–60 years (24%), and older than 60 years (13%); 3.7% of this sample did not response about their age or sex. Since data collection was organised by the show producers, and provided to us as secondary data, our research ethics committees considered the study exempt.

Identification of facial emotion expressions task

A modified version of the Identification of Emotion Expressions from Composite Faces task from the Berlin Emotion task battery (BeEmo; Wilhelm, Hildebrandt, Manske, Schacht, & Sommer, 2014) was used to measure emotion perception ability. This task was selected because, of the available BeEmo tasks, it was the most appropriate for presentation on TV, allowed for easy responding via a variety of mobile devices, and with the show's limit of five stimulus faces, allowed for the most items. This task is comparable to the other tasks as an indicator of emotion perception ability as indicated by confirmatory factor analyses including all tasks (Hildebrandt et al., 2015), with average performance normally distributed (skew = −.85; kurtosis = 1.42; Wilhelm et al., 2014; skew = −1.46; kurtosis = 1.73; Study 4 from Olderbak & Wilhelm, 2017). Given responses are influenced by the particular distractor emotion (see Supplemental Material), we selected stimulus faces that differed in their combination of target and distractor emotions.

In this task, participants are presented with a picture of a face where the top half of the face presents one emotion, and the bottom half presents a different emotion (see Wilhelm et al. [2014] and Supplemental Material for more details and the theoretical reasoning behind the task). We selected five of the original 72 photographs included in the full task with the following criteria: (1) average performance in the validation sample (.62–.70) was above guessing probability (.17), (2) each of the six basic emotions was sampled, and (3) both sexes were sampled. The persons in the photographs were Caucasian 18–30 years old (see Figure 1). The full facial expressions, coded by iMotions Emotient SDK 4.1 software (iMotions, 2016), expressed the respective emotion at 95% intensity on average. In contrast to the original version, photographs were presented in colour, and in a higher resolution (200 × 300 pixels with a 75 dpi resolution). Participants were asked to select the emotion presented in the top half of the face and in the bottom half of the face, resulting in two items per picture or 10 items in total. Respondents selected between four of the six basic emotions (anger, disgust, fear, happy, sad, surprise), with the set of emotions differing for each picture. Internal consistency was weaker than the full scale ($\alpha = .81$; Wilhelm et al., 2014; $\alpha = .95$; Study 4 from Olderbak & Wilhelm, 2017), but still acceptable ($\alpha = .52$; see Table 2 for item-level statistics and a summary of missing data).

Procedure

Over the course of the 2-hour show, 60 tasks were presented, each designed to measure a facet of the four branches of ability emotional intelligence, however, aside from the task on which the present research focuses, most of the other tasks were created for the purpose of the show and did not meet the reliability and validity standards for scientific research. The emotion composite task was presented 35 min into the show and lasted 5 min. Pictures were presented one at a time, with a break in between informing participants of the correct response for the top and

Figure 1. Presented stimuli. *Source*: permission obtained.

bottom half of the face.[1] The hosts of the show made several unrelated comments during the presentation of the pictures, which may have biased participants' responses. Pictures and response options were presented both on TV and on participants' mobile device. Participants had 20 s to answer each item.

Results

Generalisability

To evaluate the generalisability of these results, we first examined the test's validity followed by a power analysis given different estimates of test reliability.

Validity

To establish the construct validity of this test, we examined properties of the short-form solution of our measure in two external samples (Table 1). Average performance on the short-form is strongly correlated with average performance on the full version and moderately related with performance on other emotion perception tests, showing the short-form has adequate construct validity.

Power analysis

Next, we evaluated the power of this test to detect mean differences between our groups. While traditionally researchers focus on utilising multiple measures, relative to a smaller sample size, the strengths of this study are in the very large and diverse sample, relative to a more limited measurement. A simulation study showed that increased measurement error in the

Table 1. Evaluating the Identification of Emotion Expressions from Composite Faces task short-form validity through two external samples.

	Wilhelm et al. (2014)	Olderbak and Wilhelm (2017) – Study 4
Sample characteristics		
n	269	214
Age *M* (*SD*)	26.0 (5.9)	37.5 (13.5)
% Female	52%	69%
Language	German	English
Correlation with the full version	$r = .59$, $p < .001$	$r = .77$, $p < .001$
Correlation with other emotion perception tests (tasks from Wilhelm et al., 2014)	Upright Inverted: $r = .35$, $p < .001$ Visual Search: $r = .30$, $p < .001$	Upright Inverted: $r = .23$, $p < .001$

Table 2. Item-level descriptive statistics, with item-total correlations and inter-item tetrachoric correlations.

Item	Picture	Stimuli Sex	Half	Target Emotion	% Missing	*M*	*SD*	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Total
1	1	Male	Top	Anger	1.9	.575	.494										.518
2	1	Male	Bottom	Disgust	1.9	.575	.494	.779									.517
3	2	Female	Top	Sad	2.5	.303	.459	.005	.013								.305
4	2	Female	Bottom	Disgust	2.5	.303	.460	−.002	−.002	.287							.300
5	3	Male	Top	Fear	3.4	.370	.483	.111	.116	.052	.045						.395
6	3	Male	Bottom	Surprise	3.4	.367	.482	.116	.107	.058	.052	.486					.399
7	4	Female	Top	Anger	4.2	.473	.499	.186	.180	.045	.046	.077	.083				.479
8	4	Female	Bottom	Happy	4.2	.473	.499	.180	.175	.043	.044	.078	.092	.680			.478
9	5	Male	Top	Sad	5.6	.503	.500	.173	.173	.057	.058	.081	.081	.133	.128		.490
10	5	Male	Bottom	Surprise	5.6	.503	.500	.183	.180	.056	.049	.079	.081	.127	.129	.715	.490

dependent variable, due to poor reliability (e.g. through a reduced number of measures or through the use of a short-form), can be effectively minimised with a sufficiently large sample size without an impact on the estimated effect size (Hutcheon, Chiolero, & Hanley, 2010). Likewise, many have pointed out the benefits of short-form measures (e.g. Rammstedt & John, 2007). Based on the current sample size, and an average of standard deviations in performance at the group-level, averaged across groups ($SD_M = .20$), we estimated the magnitude of detectable mean differences between groups under varying estimates of test reliability given certain power assumptions.[2] As is illustrated in Figure 2, the large sample size allowed us to detect very small differences between group-level means, even assuming poor internal consistency. Under the assumption of having the lowest reliability estimate ($R = .10$), we still have sufficient power ($\alpha = .05$, $\beta = .95$) to detect very weak effect sizes ($r = .048$, $d = .097$).

Missing data

Because of how the items were presented, participants had to respond to both parts of one picture in order for their responses to be scored; however, individual pictures could be skipped completely. The majority of the sample completed every item (92%), with participants then missing data for one (3%), two (2%), three (2%), or four pictures (2%). Older participants had slightly less missing data, relative to younger participants ($r = -.02$, $p < .001$) and men had more missing data ($M = .41$ items, $SD = 1.48$) relative to women ($M = .32$, $SD = 1.31$; $t_{(100413)} = -.10.0$, $p < .001$, $r = .03$). Missing more items was associated with an average poorer performance ($r = -.19$, $p < .001$). Since the percent of missing data increases across the presentation of the test (Table 2), and since participants received feedback after each picture, we interpret this relation to reflect that some persons who, once they did poorly on a picture, dropped out and did not complete the rest of the items. Thus, this data is missing not at random. In order to maintain the representativeness of our sample, all participants have been retained.

Item-level analyses

Participants, on average, successfully identified the expressed emotion for 44% of the items ($SD = .22$; scale-level scores were average performance across all completed items. Average performance ranged from 0% to 100% and the scores were normally distributed (skew = .09, kurtosis = −.53). For each item, the correct emotion was the most often selected response, and the proportion of people selecting that response was consistently above guessing probability (.25), with no ceiling effect (Table 2).

Establishing measurement invariance

Aiming to provide a more precise estimate of performance, and to establish that emotion perception was

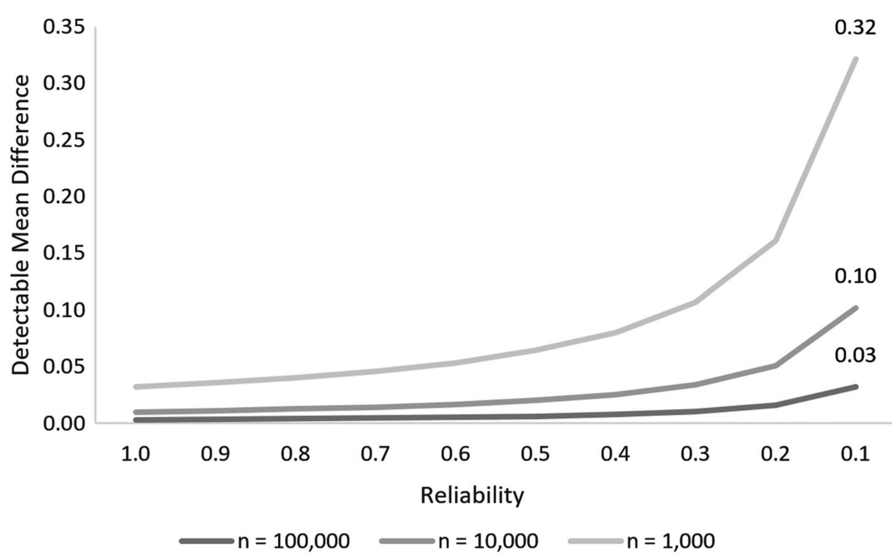

Figure 2. Detectable mean difference given varying reliability estimates ($\alpha = .05$, $\beta = .95$, $SD = .20$) with varying sample sizes.

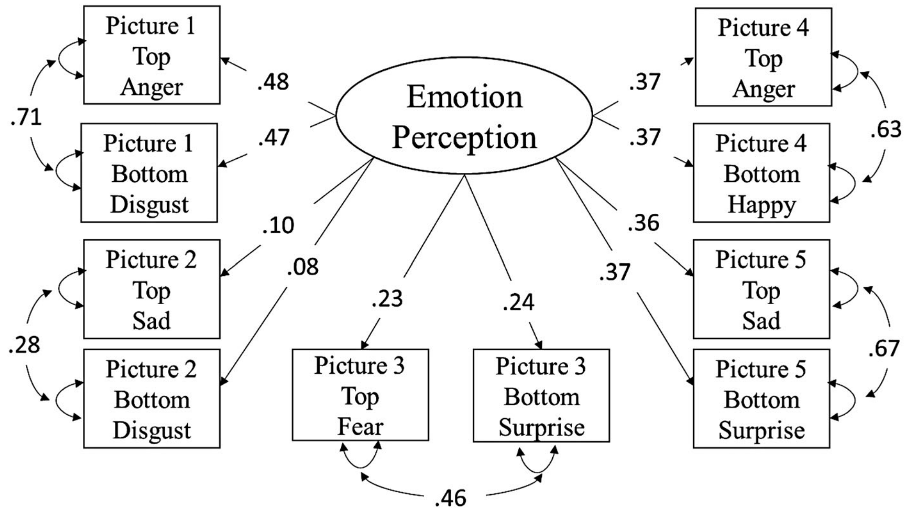

Figure 3. Final measurement model; all parameters are statistically significant ($p < .001$).

assessed invariantly for males and females and across age groups, in order to ensure the comparability of mean performance estimates between groups, average emotion perception performance was assessed with latent variable modeling (modeled with Mplus 7).[3] In a confirmatory factor analysis, we modeled a single latent variable indicated by the 10 items (latent variance set to 1) using a robust weighted least squares estimator with an adjustment of means and variances (WLSMV), which accommodates the dichotomous indicators and handles missing item-level data. According to established standards (RMSEA < .06, CFI ≥ .95; Hu & Bentler, 1999) model fit was poor ($\chi^2_{(35)} = 78771.3$, $p < .001$, RMSEA $= .147_{(.146–.148)}$, CFI = .587), but significantly improved ($\Delta\chi^2_{(5)} = 54918.7$, $p < .001$) with added covariances between the residuals of indicators from the same picture ($\chi^2_{(30)} = 549.9$, $p < .001$, RMSEA $= .013_{(.012–.014)}$, CFI = .997; Figure 3).

We confirmed the equivalence of our measurement model between males and females and across the age groups using multiple group analysis, applying the two-step approach recommended for categorical data. For each multiple group comparison, there was adequate fit for step 1, configural invariance, where we fixed scale factors to 1 and factor means to 0, while allowing item thresholds and factor loadings to vary between groups (Table 3). For step 2, strong invariance, item thresholds and factor loadings were constrained to be equal between groups, with the scale factors and factor means in one group constrained to be 1 and 0 respectively and the scale factors and factor means allowed to vary in the other groups. In step 2, model fit decreased significantly, according to the change in chi-square. However, this statistic is highly dependent on sample size and so it is untrustworthy in evaluating measurement invariance with these data. Instead, we

Table 3. Model fit from the multiple group analyses.

#	Model	χ^2	df	p	CFI	RMSEA	ΔModels	$\Delta\chi^2$	Δdf	p	ΔCFI	ΔRMSEA
Comparing males and females												
1	Configural invariance	561.5	60	<.001	.997	.013						
2	Strong invariance	1264.4	75	<.001	.994	.018	2 vs 1	536.0	15	<.001	.003	.005
Comparing age groups												
3	Configural invariance	750.0	150	<.001	.997	.014						
4	Strong invariance	2029.8	207	<.001	.991	.021	4 vs 3	1020.9	57	<.001	.006	.007
Comparing age groups by sex												
5	Configural invariance	885.4	300	<.001	.997	.014						
6	Strong invariance	2358.6	427	<.001	.990	.021	6 vs 5	1199.8	127	<.001	.007	.007

Note: Because the data were estimated with WLSMV, the $\Delta\chi^2$ is estimated through the MPlus DIFFTEST. Models 1 and 2 included two groups, models 3 and 4 included five groups, and models 5 and 6 included ten groups.

Table 4. Means of latent variables with 95% confidence intervals.

	Females and Males			Males			Females		
Age range	Lower 95% CI	Mean	Upper 95% CI	Lower 95% CI	Mean	Upper 95% CI	Lower 95% CI	Mean	Upper 95% CI
<15 years		.00		−.33	−.25	−.18		.00	
15–30 years	.39	.43	.47	.08	.14	.20	.35	.40	.46
31–45 years	.20	.24	.28	−.13	−.07	−.01	.18	.24	.29
46–60 years	−.22	−.18	−.13	−.43	−.37	−.31	−.31	−.25	−.19
>60 years	−.76	−.71	−.67	−1.02	−.95	−.88	−.83	−.79	−.70

Note: Estimates for females and males combined are from model 4, and effect sizes for females and males separately are from model 6. Means of latent variables are estimated relative to a comparison group, where the mean was set to 0 for identification purposes. The reference group for model 4 is females and males <15 years of age and for model 6 it is females <15 years of age; no confidence intervals are provided for these groups because their mean is fixed.

evaluated changes in the CFI and RMSEA, which were well below recommended cutoff criteria (ΔCFI ≥ −.010, ΔRMSEA ≥ .015; Chen, 2007), indicating measurement invariance between men and women and across age groups. Thus, the means of latent variables can be compared across sex and age groups, because the meaning of the factors is equivalent for those groups.

Performance by age and sex

Generally, performance gradients of age differences were comparable for males and females (Tables 4 and 5, Figure 4). Persons less than 15 years of age performed worse than persons 15–45 but better than persons 46 years and older. Persons 15–30 years of age performed better than all other age groups. Persons 31–45 performed worse than persons 15–30 but better than all other age groups. Persons 46–60 performed worse than persons younger than themselves, but better than persons older than themselves. Finally, persons 60 years and older performed worse relative to all other age groups.

Females consistently performed better than males (overall *M* diff = .25, 95% CI [−.26, −.23], *d* = .17, estimated from model 2), however the strength of this advantage was consistently weak in magnitude and decreased with increasing age. The advantage was largest for persons under 15 years (*d* = .14), but was almost three times smaller, *d* = .05, for persons older than 60. An examination of the effect sizes suggests this is mostly due to the large drop in ability for females between the ages of 31–45 and 46–60.

Table 5. Means of latent variables, absolute mean differences, and standardised mean difference (Cohens *d*) by sex and age.

	Comparison group				
Reference group	<15 years	15–30 years	31–45 years	46–60 years	>60 years
Females and Males					
<15 years		.12	.06	−.05	−.26
15–30 years	−.43		−.04	−.15	−.30
31–45 years	−.24	.18		−.10	−.24
46–60 years	.18	.60	.42		−.15
>60 years	.71	1.14	.95	.54	
Females only					
<15 years		.13	.07	−.08	−.33
15–30 years	−.40		−.04	−.15	−.30
31–45 years	−.24	.17		−.11	−.25
46–60 years	.25	.65	.49		−.14
>60 years	.79	1.20	1.03	.54	
Males only					
<15 years		.13	.06	−.04	−.25
15–30 years	−.39		−.06	−.15	−.33
31–45 years	−.18	.20		−.08	−.25
46–60 years	.12	.50	.30		−.18
>60 years	.70	1.09	.88	.58	

Note: Estimates for females and males combined are from model 4, and effect sizes for females and males separately are from model 6. Means of latent variables are estimated relative to a comparison group, where the mean was set to 0 for identification purposes. The reference group for model 4 is females and males <15 years of age and for model 6 it is females <15 years of age. Absolute mean difference between the reference and comparison groups are presented below the diagonal with standardised mean differences (Cohen's *d*) presented above the diagonal. 95% confidence intervals of the means of latent variables are provided in supplementary material.

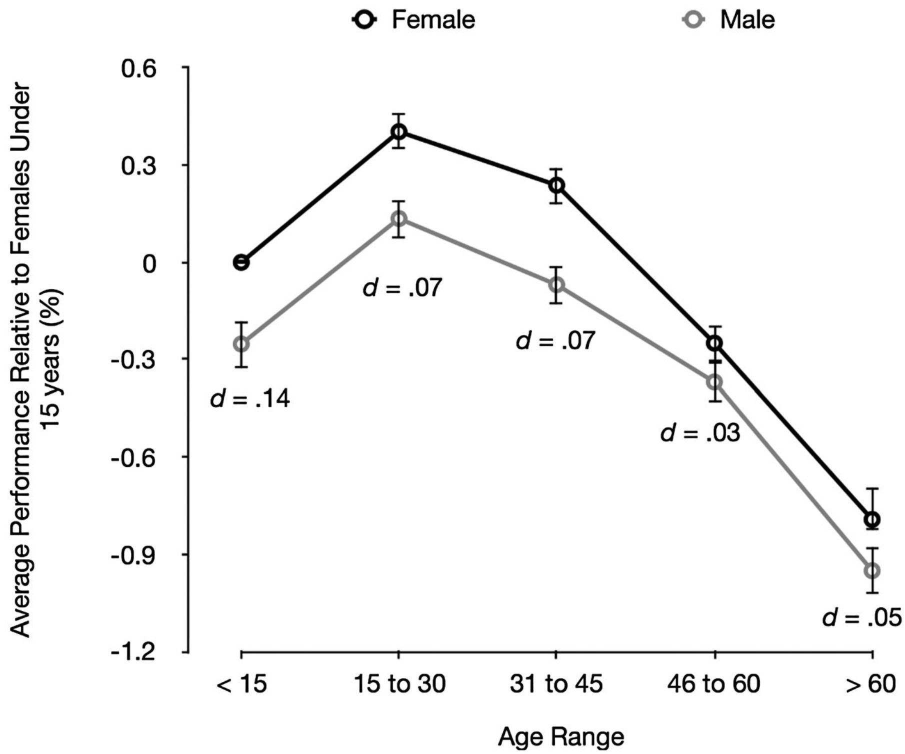

Figure 4. Latent variable mean estimates from model 6, which establishes strong measurement invariance across males and females for each age group, with 95% confidence intervals. Standardised mean differences between males and females (as Cohen's *d* estimates) are indicated in the figure, with positive scores indicating a female advantage. Absolute mean differences are as follows: <15 *M* diff = .25; 15–30 *M* diff = .27; 31–45 *M* diff = .31; 46–60 *M* diff = .12; >60 *M* diff = .16.

Discussion

Using an exceptionally large lifespan sample we found that, for both males and females, emotion perception abilities peak between the ages of 15 and 30, with poorer performance by younger adults and declining performance after the age of 30. Although females at all ages were consistently better than males at reading emotional faces, this advantage decreased with age.

The magnitude of the sex effect in emotion perception revealed in our study ($d = .17$) was comparable to meta-analytic estimates ($d = .19$; Thompson & Voyer, 2014), however our study shows for the first time that the magnitude of this difference decreases with age to become extremely small for persons over 45 ($d = .05$). That we were able to find a sex by age interaction, in contrast to the findings of Sasson et al. (2010), is most likely due to the use of latent variable modeling, which allowed a cleaner estimate of performance for each group, in addition to the increased power with the larger sample size (Figure 2). Likewise, the overall U-shaped age trajectory in our sample was comparable to that of Williams et al. (2009) and Kessels et al. (2014). However, we were able to detect declines in performance as early as 30 years, supporting the results of Williams et al. (2009), but in contrast to those of Kessels et al. (2014), who found declines starting primal at 50 years of age.

Interestingly, the age at which emotion perception abilities started to decline in our large sample was also earlier than that of the related ability to perceive neutral faces, which drops around age 60 (Hildebrandt et al., 2010). This suggests that the strong relation found between the abilities to perceive emotional versus neutral faces (e.g. Hildebrandt et al., 2015) may decrease with increasing age, which is in contrast to the differentiation-dedifferentiation hypothesis where correlations between cognitive abilities should increase with increasing age (Tucker-Drob & Salthouse, 2008). Future research should parse out the effect of declines in face perception abilities on emotion perception abilities. Likewise, research should isolate the uniqueness of sex differences in emotion perception ability from known sex differences in face perception ability (Sommer et al., 2013).

Given the strong relation between emotion perception ability and general mental ability (e.g. Hildebrandt et al., 2015), we postulate that a primary cause of the decline in emotion perception ability is the decline in general cognitive abilities (e.g. Lövdén & Lindenberger, 2005). Likewise, given strong relations between the ability to perceive emotional faces with the ability to perceive neutral faces (Hildebrandt et al., 2015), the female advantage in emotion perception ability may due to the same reasons argued for female advantage with the perception of neutral faces: a greater interest in social interactions (e.g. Kaplan, 1978).

While a multivariate measurement approach would have been preferred, we found the expected results and replicate for a unique sample what was found in better controlled settings with multivariate measures. Our approach to collect data during a popular television show has obvious limitations (e.g. restrictions on testing time, interference by television hosts or family at home, with participants viewing the stimulus material and responding through a variety of handheld devices with varying levels of technological skill), however we recruited a sample over 100 times larger than any previous study, reaching a wide range of people unlikely to participate in a typical psychological study. Through a power analysis we showed the increased measurement error due to limited measurement is drastically reduced due to the large sample size; hence, we had sufficient power to detect very small differences between groups. Together with the advanced analytical methods utilised (latent variable modeling and invariance testing of the measure), our results clearly add to the literature in terms of generalisability across the population.

With the limited number of items, the task had an uneven distribution of target emotions, along with target/distractor pairings. However, studies that utilise a multi-task approach find very weak evidence for unique and distinct emotion-specific perception abilities (e.g. anger perception ability; Hildebrandt et al., 2015; Olderbak, Mokros, Nitschke, Habermeyer, & Wilhelm, 2018; Olderbak & Wilhelm, 2017). We note, however, that there may be affect-specific trends. For example, previous studies suggest that age might be associated with poorer perception ability for negative affect emotions (i.e. anger, fear, sadness; Ruffman et al., 2008) and that older individuals may be particularly motivated to attend to positive (vs. negative) emotional stimuli (Carstensen, Fung, & Charles, 2003). Unfortunately, because these emotions were only presented on the top half of the face, we cannot statistically separate whether effect sizes associated with these emotions are due to the emotions themselves, or due to an aspect of perceiving emotion from the top half of the face. This is an issue that should be examined in future research.

Age may bring us many desirable qualities like wisdom, harmonious relationships, or even happiness, but better emotion perception is not one of them. The present research is unique in establishing the age gradient by sex in emotion perception in a very large population, showing that past 30 years old, our ability to read other people's emotion start to decline. While this study and previous research does not allow us to disentangle whether these differences are driven by developmental or generational mechanisms, our large scale investigation with a wide range of people and testing environments suggests that these effects are, none the less, largely robust.

Notes

1. Improvement over the course of the task, assessed by subtracting average performance on items 1–6 from average performance on items 7–10, was essentially unrelated with age (Men: $r = .03$, $p < .001$; Women: $r = .02$, $p < .001$), suggesting younger people did not benefit more from the verbal feedback, relative to older adults.
2. Based on a modification of equation 1.6 from Fleiss, 1986, p. 5, and incorporating the standard equation that the sample size needed is that estimated by the power analysis (N) divided by the reliability estimate (R).
3. Unbiased hit rate scoring was also inappropriate, given only 10 items were administered and there was an insufficient number of items from which to identify person-specific patterns of responding.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Carstensen, L. L., Fung, H. H., & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and Emotion*, *27*, 103–123.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 464–504.

Elfenbein, H. A., & Eisenkraft, N. (2010). The relationship between displaying and perceiving nonverbal cues of affect: A meta-analysis to solve an old mystery. *Journal of Personality and Social Psychology*, *98*, 301–318. doi:10.1037/a0017766
Fleiss, J. (1986). *The design and analysis of clinical experiments*. New York, NY: Wiley.
Gunning-Dixon, F. M., Gur, R. C., Perkins, A. C., Schroeder, L., Turner, T., Turetsky, B. I., … Gur, R. E. (2003). Age-related differences in brain activation during emotional face processing. *Neurobiology of Aging*, *24*, 285–295. doi:10.1016/S0197-4580(02)00099-4
Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*, 433–443. doi:10.1177/0956797614567339
Haxby, J. V., & Gobbini, M. I. (2011). Distributed neural systems for face perception. In *The Oxford handbook of face perception* (pp. 93–110). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199559053.013.0006
Hildebrandt, A., Herzmann, G., Sommer, W., & Wilhelm, O. (2010). Structural invariance and age-related performance differences in face cognition. *Psychology and Aging*, *25*, 794–810. doi:10.1037/a0019774
Hildebrandt, A., Sommer, W., Schacht, A., & Wilhelm, O. (2015). Perceiving and remembering emotional facial expressions — A basic facet of emotional intelligence. *Intelligence*, *50*, 52–67. doi:10.1016/j.intell.2015.02.003
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. doi:10.1080/10705519909540118
Hutcheon, J. A., Chiolero, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *BMJ*, *340*, c2289. doi:10.1136/bmj.c2289
iMotions. (2016, March). Emotient SDK 4.1 [computer software]. Copenhagen, Denmark.
Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging*, *22*, 147–159. doi:10.1037/0882-7974.22.1.147
Kaplan, H. B. (1978). Sex differences in social interest. *Journal of Individual Psychology*, *34*, 206–209.
Kessels, R. P., Montagne, B., Hendriks, A. W., Perrett, D. I., de Haan, E. H. F., & Haan, E. H. (2014). Assessment of perception of morphed facial expressions using the Emotion Recognition Task: Normative data from healthy participants aged 8–75. *Journal of Neuropsychology*, *8*, 75–93. doi:10.1111/jnp.12009
Lövdén, M., & Lindenberger, U. (2005). Development of intellectual abilities in old Age. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 203–221). Thousand Oaks, CA: Sage.
MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion*, *14*, 358–374. doi:10.1037/a0034755
Mather, M. (2016). The affective neuroscience of aging. *Annual Review of Psychology*, *67*, 213–238. doi:10.1146/annurev-psych-122414-033540
Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review*, *8*, 290–300. doi:10.1177/1754073916639667
McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*, 424–453. doi:10.1037/0033-2909.126.3.424
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
Olderbak, S. G., Mokros, A., Nitschke, J., Habermeyer, E., & Wilhelm, O. (2018). Psychopathic men: Deficits in general mental ability, not emotion perception. *Journal of Abnormal Psychology*. Advance online publication. doi:10.1037/abn0000340
Olderbak, S., & Wilhelm, O. (2017). Emotion perception and empathy: An individual differences test of relations. *Emotion*, *17*, 1092–1106. doi:10.1037/emo0000308
Park, D. C., Polk, T. A., Park, R., Minear, M., Savage, A., & Smith, M. R. (2004). From the cover: Aging reduces neural specialization in ventral visual cortex. *Proceedings of the National Academy of Sciences*, *101*, 13091–13095. doi:10.1073/pnas.0405148101
Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *Journal of Research in Personality*, *41*, 203–212.
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 & Biobehavioral Reviews*, *32*, 863–881. doi:10.1016/j.neubiorev.2008.01.001
Sasson, N. J., Pinkham, A. E., Richard, J., Hughett, P., Gur, R. E., & Gur, R. C. (2010). Controlling for response biases clarifies sex and age differences in facial affect recognition. *Journal of Nonverbal Behavior*, *34*, 207–221. doi:10.1007/s10919-010-0092-z
Sommer, W., Hildebrandt, A., Kunina-Habenicht, O., Schacht, A., & Wilhelm, O. (2013). Sex differences in face cognition. *Acta Psychologica*, *142*, 62–73. doi:10.1016/j.actpsy.2012.11.001
Sullivan, S., Campbell, A., Hutton, S. B., & Ruffman, T. (2017). What's good for the goose is not good for the gander: Age and gender differences in scanning emotion faces. *The Journals of Gerontology: Series B*, *72*, 441–447. doi:10.1093/geronb/gbv033
Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, *28*, 1164–1195. doi:10.1080/02699931.2013.875889
Tucker-Drob, E. M., & Salthouse, T. A. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and Aging*, *23*, 453–460. doi:10.1037/0882-7974.23.2.453
Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in Psychology*, *5*, 1–23. doi:10.3389/fpsyg.2014.00404
Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology*, *31*, 257–277. doi:10.1080/13803390802255635