



Published in final edited form as:

Cogn Emot. 2019 March ; 33(2): 245–257. doi:10.1080/02699931.2018.1445981.

Emotion identification across adulthood using the Dynamic FACES database of emotional expressions in younger, middle aged, and older adults

Catherine A. C. Holland^{1,*}, Natalie C. Ebner^{2,3}, Tian Lin², and Gregory R. Samanez-Larkin^{1,4,*}

¹Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven CT 06520.

²Department of Psychology, University of Florida, P.O. Box 112250 Gainesville, FL 32611-2250

³Department of Aging and Geriatric Research, Institute on Aging, University of Florida.

⁴Department of Psychology and Neuroscience, Duke University, Box 90999, Durham, NC 27708.

Abstract

Facial stimuli are widely used in behavioral and brain science research to investigate emotional facial processing. However, some studies have demonstrated that dynamic expressions elicit stronger emotional responses compared to static images. To address the need for more ecologically valid and powerful facial emotional stimuli, we created Dynamic FACES, a database of morphed videos ($n = 1,026$) from younger, middle-aged, and older adults displaying naturalistic emotional facial expressions (neutrality, sadness, disgust, fear, anger, happiness). To assess adult age differences in emotion identification of dynamic stimuli and to provide normative ratings for this modified set of stimuli, healthy adults ($n = 1822$, age range 18–86 years) categorized for each video the emotional expression displayed, rated the expression distinctiveness, estimated the age of the face model, and rated the naturalness of the expression. We found few age differences in emotion identification when using dynamic stimuli. Only for angry faces did older adults show lower levels of identification accuracy than younger adults. Further, older adults outperformed middle-aged adults' in identification of sadness. The use of dynamic facial emotional stimuli has previously been limited, but Dynamic FACES provides a large database of high-resolution naturalistic, dynamic expressions across adulthood. Information on using Dynamic FACES for research purposes can be found at <http://faces.mpib-berlin.mpg.de>.

Keywords

emotion; faces; adulthood; aging; dynamic expressions

*Correspondence concerning this article should be addressed to Catherine Holland (catherineholland@fas.harvard.edu) or Gregory Samanez-Larkin (g.samanezlarkin@duke.edu).

Introduction

Facial expressions of emotion are unique stimuli with significant social and biological relevance. They are used in a wide range of studies across various fields of behavioral and brain science to investigate attention, memory, social reasoning, perception, learning, decision-making, motivation, and emotion regulation (Davis, 1995; Eimer et al., 2003; Elfenbein, 2006; Fotios et al., 2015; Tran et al., 2013).

As humans develop so do their facial-processing abilities (Pascalis et al., 2011). Across the adult life span, facial-processing abilities appear to decline in that facial emotional expression identification becomes less accurate in older age (Ruffman et al., 2008). Additionally, older adults preferentially remember positive relative to negative faces (Mather et al., 2003), and older adults attend more to neutral than negative faces (Isaacowitz et al., 2006).

Of note, in many of the previous studies facial expression stimuli have been exclusively used from younger adults and/or presented as static images. Over the past several years, a number of dynamic emotional expressions stimulus sets have emerged, but nearly all of them are composed of expressions from mostly younger adults and/or include normative rating data from mostly young adults (Krumhuber et al., 2017). Very few collections of facial stimuli for use in research contain images of individuals that vary in age across the adult life span. Given that variations in the characteristics of facial stimuli, such as age, gender, and expression differentially influence how a face is evaluated and remembered (Ebner, 2008; Golby et al., 2001; Isaacowitz et al., 2006; Mather et al., 2003; Pascalis et al., 1995; Ruffman et al., 2008), there is a need for more naturalistic facial expression stimuli across a broad range of ages in women and men.

Moreover, there is evidence for an own-age bias in facial recognition wherein children, middle-aged, and older adults are more likely to recognize faces of their own compared to another age group (Rhodes et al., 2012). As a result, it is possible that older adults have been at a disadvantage in prior studies due to a lack of appropriate stimuli. The FACES Lifespan Database was created (Ebner et al., 2010) to provide stimulus material that would overcome this disadvantage. FACES is a database which includes six categorical facial emotional expressions (neutral, happy, angry, disgusted, fearful, sad) from younger, middle-aged, and older women and men.

The aim of the Dynamic FACES project was to extend the existing FACES database by creating and validating a set of dynamic facial stimulus videos based on the original FACES images. Recent research has begun to demonstrate the significance of dynamic information in facial processing. Indeed, brain regions such as the amygdala and fusiform face area show greater responses to dynamic than static emotion expressions (Freyd, 1987; LaBar et al., 2003). This finding is particularly pronounced for negative emotions such as fear (LaBar et al., 2003). Studies have further demonstrated that dynamic properties affect the perception of facial expressions such that movement enhances emotion identification accuracy and increases perceptions of emotional intensity (Kamachi et al., 2013). Relatedly, dynamic stimuli may also be useful in research with clinical populations, such as individuals with

autism spectrum disorder, whose emotion identification is partially mediated by the dynamics of emotional expressions (Blair et al., 2001; Gepner et al., 2001). As real-life facial interactions are typically dynamic in nature, dynamic research stimuli constitute more naturalistic stimuli than static images of facial expressions. Thus, use of dynamic facial stimuli has the potential to advance understanding across a range of individual differences in facial emotion processing from age differences to clinical comparisons.

The primary goal of the study was to create a large set of controlled and high-resolution videos that powerfully communicated emotion. Although other collections of dynamic facial stimuli exist (Krumhuber et al., 2017), the Dynamic FACES database is unique in that it is comprised of the largest and most evenly distributed sample of younger, middle-aged, and older women and men, each displaying six facial expressions (*neutrality, sadness, disgust, fear, anger, and happiness*). Dynamic stimulus videos were created by morphing neutral facial expressions to emotion expressions. Morphing allows the researcher to control the duration of the emotion presentation, a feature that will be useful for fMRI researchers. To provide normative ratings of these adapted stimuli, we collected rating data from younger, middle-aged, and older women and men on facial expression identification, expression distinctiveness, perceived age, and naturalness of the facial expression displayed for each video. Ratings were then analyzed for age-group differences to examine effects of age on emotion identification using dynamic stimuli (e.g., Ebner, 2008; Ebner et al., 2010; Issacowitz et al., 2007; Ruffman et al., 2008). That is, the primary goal of the study was to create a dynamic stimulus set from a broad age range of face models; the investigation of age differences in emotion identification was a secondary goal.

Method

Selection of Images from the FACES Lifespan Database

Face Models.—The static images used for Dynamic FACES were taken from the original FACES database created at the Max Planck Institute for Human Development in Berlin, Germany (Ebner et al., 2010). The FACES database is comprised of 58 healthy Caucasian women and men ($N = 171$). These individuals were recruited through a model agency (“Agentur Wanted,” Berlin) and were selected for their “average” appearances and non-prominent eye-catching facial features. Models were asked to wear a standardized gray shirt and remove all jewelry, make-up, and glasses prior to the photo-shoot (Figure 1). The original expressions in the static images were posed. Each model underwent a face training exercise including (1) an *emotion induction* phase, aimed at triggering the spontaneous experience of the emotion; (2) a *personal experience* phase, aimed at inducing the emotion by imagining personal events that had elicited the emotion, and (3) a *controlled expression* phase, in which the models were trained to move and hold specific facial muscles to optimally display each emotion. Before photographs were taken, models were instructed to display each facial expression as intensively as possible but in a natural-looking way (for more information on the emotion induction tasks see Ebner et al., 2010). Each model in the FACES database is represented by two sets of six facial expressions (*neutrality, sadness, disgust, fear, anger, and happiness*) totaling 2,052 images. For more information on the development and validation of the FACES database see Ebner et al. (2010).

The aim of the picture selection process for Dynamic FACES was to identify the one image per expression per face model that best represented each of the six facial expressions. These were selected based on the higher mean percent accuracy of emotion expression identification by all raters for each set of emotions as assessed in Ebner et al. (2010). This selection process resulted in a total of 1,197 images to be morphed for the video stimuli creation (Figure 1). Although only one image per expression was selected, both neutral images per face model in the original FACES database were needed to create the neutral expression morphs (see description below). Table 1 summarizes the final age and gender distributions for the Dynamic FACES database.

Creation of Dynamic FACES

The dynamic facial stimulus videos were created between July 2015 and April 2016 by seven Yale undergraduate research assistants using the program Morph Age for OS X (Figure 2). Each research assistant completed a lab-internal training session led by C.A.C.H. and consulted a manual with video tutorials made by C.A.C.H. specifically for morphing the images (manual available on request).

The Morph Age program works by allowing a user to define curves on one or more images. Any change made on the curves is reflected on the resulting image through a corresponding morphing effect. Each video was created by starting with a model's highest mean percent accuracy neutral image and was then morphed into one of the six highest mean percent accuracy target facial expressions. Both neutral expression sets from FACES were used to create the neutral expression videos. In this case, the morphs started from the neutral image in FACES *set a* and ended in the neutral image in FACES *set b*. All videos and morphage files were reviewed for quality (including natural development of the expression and morphing artifacts such as jumping eyebrows/wrinkles or overlapping lips and teeth) by at least two research assistants. If one research assistant thought the video could be improved, the two research assistants would correct the morph together.

Each of the final 1,026 videos is exactly two seconds long (one second for the morph to occur, followed by one second for the expression to statically persist). The duration of the morph was selected so that the videos would appear naturalistic and did not evolve so fast that they were surprising (based on subjective evaluations of morphs that ranged from 200ms to 2000ms). Holding the expression for one second following the morph was also based on initial testing to maximize the naturalness of the emotion expression in the videos. These video durations also conveniently coincide with the typical repetition time for a whole-brain acquisition in functional neuroimaging experiments (making them particularly well-suited for research using functional magnetic resonance imaging). The final Dynamic FACES videos were stored in .mp4 format with a frame rate of 30 fps and resolution of 384×480 pixels. They were re-sized from a .mov format with dimensions $1,280 \times 1,600$. The original videos and morphage files (from which the videos were generated) have also been archived and are available upon request. These files can be used by experimenters seeking a different video length, duration of emotion appearance, and/or frame-rate. The manual for altering these dimensions is also available upon request.

For easy reference, all videos for a given model were saved with their FACES personal code and the first letter of their respective age-group, gender, and neutral-to-target facial expression (ex. 006_m_f_n_d.mp4 is the video of a middle-aged female being morphed from neutral to disgust).

Validation of Dynamic FACES

To validate the Dynamic FACES database and provide video-specific information as a reference for researchers, each of the 1,026 videos were rated in terms of facial expression displayed, expression distinctiveness, perceived age, and naturalness of the morph. We collected ratings of expression distinctiveness as a particularly relevant feature for memory research (Watier & Collin, 2012). Perceived age was included as a rating dimension given the broad age range of face models in the videos. This is one of the few databases that includes faces of different ages, and previous research has shown that perceived age can vary across facial expressions (Voelkle et al., 2012).

A total of 2,208 raters took part in this study. 358 participants (age mean = 38.77, SD = 15.34) were excluded for expression emotion identification accuracy below 33% and 28 participants were excluded for typing in nonsense/non-word responses, resulting in 1,822 raters included in the final analysis (age mean = 48.57, SD = 16.14). Mathematically chance accuracy with six response options is 16.67%. While this threshold would remove completely random responders, it would not remove raters with inconsistent engagement over the course of the task. When visualizing the distribution of accuracy rates, there appeared to be a break point below 33% with very few raters between 33% and 60% accuracy. The data were collected online so it is difficult to know exactly why accuracy was so low for a given excluded individual. However, all of the individuals below 33% accuracy completed the entire study very quickly and nearly all showed low accuracy rates for happiness (which was nearly at perfect accuracy for the included raters). Table 2 summarizes the number and demographic information of the final validation sample, broken down by age, gender, and level of education. The final rater sample was comprised of 639 younger ($M=29.8$ years, $SD = 5.9$, age range: 18–39), 688 middle-aged ($M = 50.6$ years, $SD = 5.8$, age range: 40–59), and 489 older ($M = 68.7$ years, $SD = 5.2$, age range: 60–86) raters. Participants were recruited via Qualtrics Panels and were native English speakers.

Prior to the start of the *Face-Rating Task* (described below), participants were asked a series of demographic questions including age, gender, race, level of education, and socioeconomic status. Individuals were also screened for having experienced neurological and psychological health issues. Before the first rating trial, a test video was used to screen out participants who could not play the videos. If participants indicated they could not view the test video then the experiment ended; they never rated any actual stimuli. Individuals who could not play the Dynamic FACES videos due to technical difficulties through Qualtrics and/or their web browsing software were excluded from the final analysis. Data collection was conducted in two waves after discovering that a subset of videos needed editing (based on low naturalness ratings or low emotion identification accuracy) or did not get sufficient ratings in the first wave due to the random selection of videos for individual raters. These videos were edited and recreated by one of the authors (C.A.C.H.) to improve their

naturalness and overall video quality. New rating data were collected for those revised videos (see Methods below). All data were combined across waves for analysis.

The *Face-Rating Task* started with written instructions and a practice video (that was not part of the final set of stimuli). Subsequent videos were presented one at a time, in the left top of the computer screen, with a series of four questions. Participants were instructed to play the video and make a series of judgments (in fixed order) about identifying the emotion in the video (“Which facial expression does this person primarily show?”; response options, in fixed order: *neutral, anger, disgust, fear, happiness, sadness*), distinctiveness of the emotion (“How pronounced is this specific facial expression?”; response options, in fixed order: *not very pronounced, moderately pronounced, very pronounced*), age of the individual (“How old is this person?”; input number between 0–100 years), and naturalness of the expression appearance (“How natural does this expression appear?”; response options, in fixed order: *very unnatural, unnatural, somewhat unnatural, somewhat natural, natural, very natural*). The video watching and rating were self-paced by the participants. Videos were presented one at a time for as long as it took the subject to respond to the question. On average it took subjects <15 minutes to complete the study. There were two waves of data collection. In the first wave of data collection, participants viewed 25 randomly selected videos each.

A second wave of data collection (due to technical difficulties and design limitations of the first wave of data collection as mentioned above), used the same study design listed above and was conducted for a selection of 59 videos in a new sample of participants. These videos were selected for having a low naturalness rating (below 3 on the naturalness scale) in the first wave, low emotion identification accuracy, or having less than 15 raters. Videos were edited to improve quality and re-rendered for the second wave data collection. Our analysis does not include the original (first wave) ratings for these videos. In the second wave of data collection, participants viewed 15 videos randomly selected from the re-rendered videos.

The raw rating data and our analysis code are publicly available on OSF (<https://osf.io/suu3f/>).

Results

Video-Specific Data

Each video was rated by a minimum of 16 and maximum of 58 ($M = 35$, $SD = 11.2$) raters. For each video the number of raters, mean and standard deviation for correct emotion identification, emotion distinctiveness, and naturalness were computed. This information is available as supplemental material (AppendixVideoRatings.csv).

Overall, percentages of correct emotion expression identification were relatively high and comparable to ratings of the original, static FACES images (Ebner et al., 2010). In particular, in the total sample of raters, on average, 84.9% ($SD = 35.7\%$) of the neutral faces, 97.7% ($SD = 14.6\%$) of the happy faces, 73.6% ($SD = 44.0\%$) of the angry faces, 88.2% ($SD = 32.2\%$) of the disgusted faces, 89.1% ($SD = 31.1\%$) of the fearful faces, and 72.3% ($SD = 44.7\%$) of the sad faces were correctly identified.

Emotion Identification as a Function of Rater Age, Face Age, and Face Emotion Expression

Some research suggests that individuals differ in the ability to identify facial expressions as a function of their own age and the age of the face being rated (Rhodes & Anastasi, 2012). Thus we, investigated the effects of rater age (younger, middle-age, older), face age (younger, middle-aged, older), and face emotion expression (neutral, happy, angry, disgusted, fearful, sad) on accuracy of the identification of the newly created dynamic stimuli. Tables 1 and 2 show the exact age ranges for raters and face models.

We used multilevel logistic regression to examine the main effects and all possible interactions on emotion identification accuracy (0=incorrect, 1=correct). We used cluster estimations for standard errors to accommodate for the data structure of ratings nested within raters. To evaluate the significance of the effects we ran the full model and then evaluated whether model fit significantly worsened when removing an effect using Wald statistics. We did this in two steps. The first-step regression model contained the three main effects (i.e., basic model), and we removed each main effect one at a time to evaluate whether the model fit decreased significantly. The second-step regression model included all main and interaction effects (i.e., full model), and we removed each interaction effect one at a time to evaluate whether the model fit decreased significantly.

Basic model.—Analysis of the basic model (containing the three main effects) showed that the main effect of rater age was significant ($\chi^2(2) = 15.23, p < .001$). That is, when collapsing across face age and face emotion, emotion identification accuracy was higher for younger and middle-aged raters than older raters (Figure 3a). Although this main effect was statistically significant, all of the age-group differences were very small (rightmost column in Table 3).

Further, the main effect of face age was significant ($\chi^2(2) = 334.22, p < .001$). That is, when collapsing across rater age and face emotion, emotion identification was more likely correct for younger faces than middle-aged faces and older faces and for middle-aged than older faces (Figure 3b and rightmost column in Table 4).

In addition, the main effect of face emotion was significant ($\chi^2(5) = 1163.11, p < .001$). That is, when collapsing across rater age and face age, compared to neutral faces, happy, disgusted, and fearful faces were more likely to be accurately identified while angry and sad faces were less likely to be accurately identified (Figure 3c and bottom row of Table 3 or 4). In addition, emotion identification accuracy was higher for happy than disgusted or fearful faces and was also higher for angry than sad faces.

Full model.—Analysis of the full model (containing all main effects and interactions) showed that the significant main effects were qualified by two significant two-way interactions. The two-way interaction between rater age and face emotion was significant ($\chi^2(10) = 30.04, p < .001$). In particular, there were no differences between younger, middle-aged, and older raters for neutral, happy, disgusted, or fearful faces but the age groups differed for angry and sad faces (Figure 3d and Table 3). Younger and middle-aged raters were more accurate than older raters in expression identification of angry faces.

Further, older and middle-aged raters showed higher identification accuracy than younger raters for sad faces (Figure 3d and Table 3). Confusion matrices (i.e., misidentification of one emotion for another) are presented by rater age and face emotion expression in the Supplementary Figure 1. As shown, the angry expressions were most frequently misidentified as expressions of disgust across all age groups.

There also was a significant two-way interaction between face age and face emotion ($\chi^2(10) = 32.18, p < .001$) showing that the size of the face age effect on emotion identification accuracy (i.e., younger > middle-aged > older) varied across emotion categories (Figure 3e and Table 4). In particular, the age-group differences were smaller for happy and fearful faces relative to neutral, angry, and sad faces (Figure 3e and Table 4).

The two-way interaction between rater age and face age ($\chi^2(4) = 5.55, p = .24$) and the three-way interaction ($\chi^2(20) = 25.80, p = .17$) were not significant.

Perceived Age as a Function of Rater Age, Face Age, and Face Emotion Expression

On average, younger faces were rated as 27.8 years old ($SD = 6.9, SEM = .06$; age range 18 to 40), middle-aged faces as 48.2 years old ($SD = 9.5, SEM = 0.08$; age range 40 to 60), and older faces as 66.3 years old ($SD = 8.6, SEM = .07$; age range 60 to 100). The actual average age was 24.2 years ($SD = 3.4$; age range 19 to 31) for younger face models, 49.0 years ($SD = 3.9$; age range 39 to 55) for middle-aged face models, and 73.2 years ($SD = 2.8$; age range 69 to 80) for older face models.

As in the analyses above for emotion identification, we used multilevel logistic regression to examine the main effects of rater age, face age, and face emotion expression and all possible interactions on age perception accuracy (0=incorrect, 1=correct). Ages were reported by raters as whole numbers but were scored as correct or incorrect based on whether the age was within the correct age category (younger, middle, old). We evaluated the significance of main effects and interactions as reported above.

Basic model.—Analysis on the basic model showed that the main effect of rater age was significant ($\chi^2(2) = 30.39, p < .001$). That is, when collapsing across face age and face emotion, middle-aged raters showed higher accuracy in age perception than both younger and older raters. However, the differences between the age groups were very small (Supplementary Figure 2).

Further, the main effect of face age was significant ($\chi^2(2) = 1050.44, p < .001$). That is, when collapsing across rater age and face emotion, accuracy for perceived age estimation was higher for younger than both middle-aged and older faces (Supplementary Figure 2b and Supplementary Table 1). Age perception accuracy was slightly higher for older compared to middle-aged faces but this effect was very small.

In addition, the main effect of face emotion was significant ($\chi^2(5) = 11.84, p = .037$). However, the differences in age perception accuracy between emotion categories were very small (Supplementary Figure 2c and Supplementary Table 1).

Full model.—Analysis of the full model showed that the significant main effects were qualified by a significant two-way interaction between face age and face emotion ($\chi^2(10) = 21.55, p = .018$). However, the differences between younger, middle-aged, and older faces within face emotions were all very small (Supplementary Figure 2d and Supplementary Table 1). The two-way interaction between rater age and face age ($\chi^2(4) = 2.61, p = .63$), the two-way interaction between rater age and face emotion ($\chi^2(10) = 10.63, p < .39$), and the three-way interaction ($\chi^2(20) = 20.59, p = .42$) were not significant.

Naturalness Ratings as a Function of Rater Age, Face Age, and Face Emotion Expression

The videos were rated as reasonably natural across raters and faces (4 on the rating scale = “somewhat natural”). In particular, on average, younger faces were rated as 3.99 on the naturalness scale ($SD = 1.3, SEM = .01$; age range 18 to 40), middle-aged faces as 3.94 ($SD = 1.3, SEM = 0.01$; age range 40 to 60), and older faces as 4.01 ($SD = 1.29, SEM = .01$; age range 60 to 100). In the total sample of raters, on average, neutral faces were rated as 4.42 ($SD = 1.04, SEM = 0.01$), happy faces as 4.39 ($SD = 1.19, SEM = 0.01$), angry faces as 3.69 ($SD = 1.39, SEM = 0.01$), disgusted faces as 3.73 ($SD = 1.39, SEM = 0.01$), fearful faces as 3.65 ($SD = 1.39, SEM = 0.01$), and sad faces as 3.97 ($SD = 1.28, SEM = 0.01$).

A similar approach as used above was adopted to examine effects of rater age, face age, and face emotions on naturalness ratings. However, given the continuous nature of the dependent variable, we used multilevel linear regression; naturalness ratings of the expression varied from 1 (very unnatural) to 6 (very natural). Determination of the significance of main effects and interactions followed the same approach as described above.

Basic model.—Analysis of the basic model showed that the main effect of rater age was not significant ($\chi^2(2) = 5.41, p = .067$) (Supplementary Figure 3a). The main effect of face age was significant ($\chi^2(2) = 26.26, p < .001$). That is, when collapsing across all raters and face emotions, videos of younger and older faces were rated as more natural than those of middle-aged faces. However, the differences were quite small (Cohen $f^2 < .001$; see Supplementary Figure 3b). The main effect of face emotion was also significant ($\chi^2(5) = 1243.16, p < .001$, Cohen $f^2 = .08$). That is, when collapsing across all raters and all faces, neutral and happy expressions were rated as more natural compared to the other emotions while the naturalness ratings between neutral and happy faces were not different from each other (Supplementary Figure 3c). In addition, the naturalness ratings for sad faces were significantly higher than those for angry, disgusted, and fearful faces.

Full model.—Analysis of the full model showed that the significant main effect of face emotion was qualified by a significant two-way interaction between rater age and face emotion ($\chi^2(10) = 32.69, p < .001$). However, the differences between younger, middle-aged, and older raters across face emotion categories were all very small (Cohen $f^2 = .003$; Supplementary Figure 3d and Supplementary Table 2). In addition, the significant main effects of face age and face emotion were further qualified by a significant two-way interaction ($\chi^2(10) = 49.50, p < .001$). However, the differences between younger, middle-aged, and older faces across emotion categories were all quite small (Cohen $f^2 = .003$; Supplementary Figure 3e and Supplementary Table 3). Neither the two-way interaction

between rater age and face age ($\chi^2(4) = 1.54, p = .82$) nor the three-way interaction ($\chi^2(20) = 26.94, p = .14$) was significant.

Discussion

We examined adult age differences in emotion identification using a newly developed database of experimental stimuli, the Dynamic FACES database. The primary goal of the study was to provide normative data for highly standardized dynamic facial stimuli depicting face models from a wide adult age range. Our results show high percentages of correct facial expression identification and reasonable naturalness ratings (especially given that expressions were posed in a laboratory) for these newly developed dynamic stimuli. These ratings were comparable to those for the original static images from the FACES database (Ebner et al., 2010) and were similar to ratings from other facial databases (Goeleven et al., 2008; Tottenham et al., 2009).

Currently adult age differences in evaluations of dynamic facial expressions depicting different ages of faces and emotion categories is understudied. Our study contributes to this literature with various novel findings as discussed next. The statistically significant main effect of rater age on emotion identification in the videos was very small (i.e., differences in accuracy between age groups of less than 2.5%). Further, this small main effect was qualified by a rater age by face emotion interaction. Interestingly, in contrast to earlier studies, older adults performed as well as younger and middle-aged adults in identification of neutral, happy, disgusted, and fearful faces. Although older relative to younger and middle-aged adults showed lower levels of identification accuracy for angry faces, older adults showed similar accuracy to younger adults and outperformed middle-aged adults' in identification of sadness. These findings are consistent with research suggesting that a discrete emotions approach (rather than simply positive vs negative valuation) better characterizes adult age differences in emotion processing (Kunzmann et al., 2014) especially for the processing of anger and sadness (Kunzmann et al., 2016).

Overall, the age differences were similar to, and potentially smaller than, previously reported findings using the static FACES (Ebner et al., 2010). In the present context, a direct statistical comparison of age effects and interactions between previously collected static data and the current dynamic data was not possible given the many differences between studies (e.g., number of stimuli rated per participant, German vs American raters, in-lab vs online administration). However, throughout the manuscript, we have provided estimates of effect sizes where possible.

We show here, with a very large sample of ratings, that older adults perform as well as or better than younger adults regarding accurate facial expression identification for nearly every emotion category (except anger). Dynamic images provide more information than static images and also are more naturalistic. It is possible that the increased salience of these images due to the dynamics improves performance relative to processing static images for older adults. However, it is also possible that the sample of older raters in the present study was slightly healthier than average. We did not collect health data from participants so we cannot rule out this possibility.

Other work has shown that simplifying the response options (Mienaltowski et al., 2013; Ortega, 2010) and adding context to the presentation of static faces (Noh & Isaacowitz, 2013) has been demonstrated to improve emotion identification especially in older adults. To maximize comparability of rating data with the previous data collected using the static FACES, however, we used the same fixed choice response task. The present data suggest very minimal age differences, but it is certainly possible that a different response format or a context manipulation might boost performance in older adults relative to younger adults even further.

Across all raters, happiness was the emotion with the highest emotion identification accuracy consistent with previous research (Ruffman et al., 2008). This is likely a result of happiness being the only positively-valenced emotion in the expressions investigated in our study. Happiness and neutral videos were also rated as more natural compared to the other categories followed by sadness and then anger, disgust, and fear. In contrast to previous findings, sadness and anger appeared more difficult to identify than disgust (Ebner et al., 2010; Ruffman, 2008).

Consistent with previous findings, although the effect was relatively small, emotion identification was more difficult for middle-aged and older faces than for younger faces. The accuracy of age perceptions were also much lower for middle-aged and older compared to younger faces. However, there were very minor differences between rater age groups in the accuracy of age perceptions of the faces in the videos. Consistent with previous results using the static FACES (Ebner et al., 2010), raters perceived the younger faces as older than they actually were and older faces as younger than they actually were. However, contrary to prior research we did not observe variation in age perception across emotion categories (Voelkle et al., 2012).

The major advantage of the Dynamic FACES over the static images of the original FACES database (or other similar resources) is in the dynamics. The expressions appear contingent; the videos give the viewer the sense that the individual displayed is reacting to them. These dynamic stimuli will likely be more salient when used in laboratory studies and constitute a more naturalistic representation of face emotion expressions in adults of different ages. In fact, prior research using similar morphing procedures with a small set of lower resolution images showed that dynamic compared to static stimuli significantly enhanced functional neural signals (LaBar et al., 2003). Not only are the stimuli dynamic but the static images from which they were created were high resolution resulting in very naturalistic dynamic expressions. Although synthetically morphed, we expect these newly created videos to be as powerful as dynamic stimuli used in the few existing previous studies. It is possible that some of the previously reported age differences in the neural processing of emotional facial stimuli (Samanez-Larkin & Carstensen, 2011) may be reduced when using these newly developed standardized, high-quality dynamic stimuli.

In addition to the many strengths, the present study has several limitations. The videos were created from morphed still images rather than recording of the naturalistic dynamics of expressions. However, one advantage of these morphed videos is that the expression timing is controlled across categories and that novel dynamic stimuli can be created (e.g., partial

and full expression videos with matched frame rates and timing). Although naturalness ratings and identification accuracy were relatively high, it is possible that completely natural dynamic videos would be perceived as even more naturalistic and more representative of a specific emotion expression.

An additional limitation of this stimulus set is that we did not collect test-retest data to assess the reliability of ratings within subjects. In this study, we examined consistency of ratings between raters as our measure of rating reliability. There is some evidence for inter-rater reliability as reflected in the relatively narrow standard deviation of ratings for each video (presented in the appendix). This study includes the largest number of raters of any life-span facial expression stimulus set (N=1822). Although collecting data from a large sample of raters provides an estimate of the consistency with which these stimuli will be recognized across a random sample of adults, it does not provide information about how individuals may perceive these stimuli differently after multiple exposures. That is a major limitation of our rating data.

Limiting the generalizability of the current findings and the applicability of the facial database somewhat is that all stimulus images are Caucasian. People are subject to the own-race effect when evaluating faces. That is, they are better at recognizing faces and emotions from their own race than from other races (Meissner et al., 2001; OToole et al., 1994; Walker et al., 2003). With this in mind, it is possible that the Dynamic FACES database might be best-suited for research that does not target race/ethnicity effects. Note that the raters in the present study were not all Caucasian, so race effects of raters could be explored in our publicly shared data. We had no hypotheses about race effects so did not examine it here.

Although there are over 1,000 stimulus videos in this set, the total number of available videos in Dynamic FACES for some studies may still be too small. Additionally, online data collection does not allow for control of visual angle, so the viewed size of the stimuli may have greatly varied across raters. It is also possible that visual ratings were completed with a combination of directed and averted gaze; this cannot be controlled or assessed well with online data collection. Also, the stimuli do not cover the whole life span (i.e., children, adolescents) and only basic emotions are expressed with only a single positive emotion. The original FACES database did not include surprise, so we were not able to include it here. Future databases of stimuli could include representation of more complex emotions such as pride, love, shame, and contempt, expressed not only in the face but also via voices and body postures (Banziger et al., 2012).

In sum, the Dynamic FACES database is unique in that it provides a set of high-resolution dynamic emotional stimuli that systematically vary by gender, age, and emotion for which normative rating data were collected from adults of all ages. It is comprised of 1,026 younger, middle-aged, and older female and male emotion expression videos. The videos were created using static images of the 171 face models from the FACES database (Ebner et al., 2010) available for scientific use (www.esdoc.org/JSPWiki/en/Faces). The newly created videos will be made freely available for use in the scientific research community. It will be included with the original FACES database as an eSciDoc service of the Max Planck Digital Library (MPDL; www.esdoc.org/JSPWiki/en/Faces). Functions of this online

service will include browsing, searching, viewing, selecting, and exporting of videos. Videos will be available for download in the resized resolution of 384 X 480 pixels as .mp4 files or in their original size of 1280 × 1600 as .mov files. A subset of the Dynamic FACES database will be previewed on the MPDL. To obtain access for the whole database researchers will need to register via the FACES technical agent (see <http://faces.mpib-berlin.mpg.de> for details). The morphage files from which the videos were generated will not be available after registration. In order to prevent circulation of modified Dynamic FACES videos, researchers seeking to modify videos via the original morphage files will need to submit a request to the FACES technical agent.

In creating and validating the Dynamic FACES database and making it freely available to the research community, we hope to provide a new resource that addresses the lack of dynamic facial stimuli in current research. The database was primarily designed to answer basic research questions on developmental and adult-age differences in emotional facial processing. However, the newly created images will also be useful for clinical researchers who work with adults of various ages.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Catherine Holland is now at Harvard University and Gregory Samanez-Larkin is now at Duke University. This research was conducted by the Motivated Cognition and Aging Brain Lab at Yale University. Gregory Samanez-Larkin was supported by a Pathway to Independence Award from the National Institute on Aging (R00-AG042596). The authors thank undergraduate research assistants Min Kwon, Nafessa Abuwala, Galen McAllister, Emil Beckford, Cristian Hernandez, and Yonas Takele at Yale University for their assistance in creating the face morphs. We also thank Mark Hamilton for computer programming help. All data and STATA analysis code are publicly available on OSF: <https://osf.io/suu3f/>

References

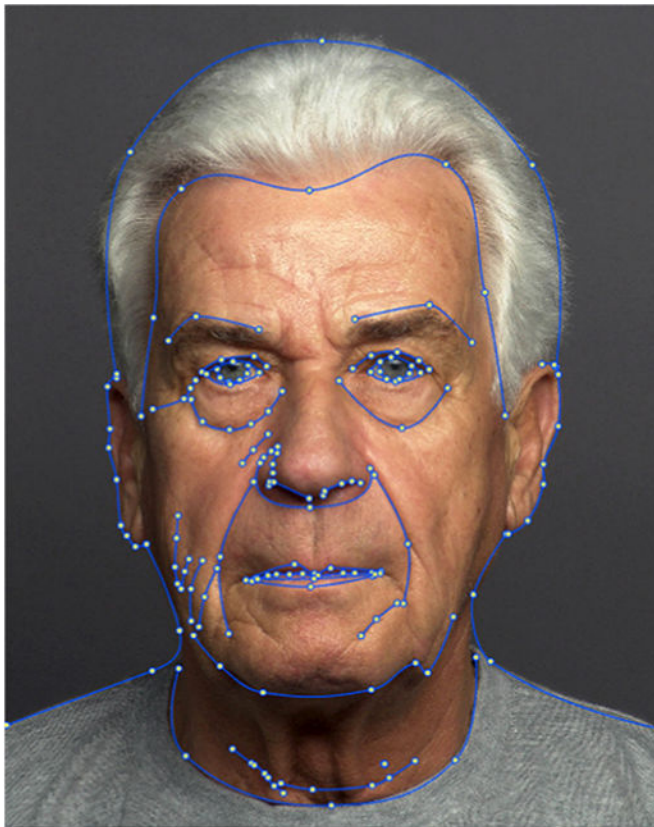
- Banziger T, Mortillaro M, & Scherer KR (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5), 1161–1179. [PubMed: 22081890]
- Blair RJR, Colledge E, Murray L, & Mitchell DGV (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of abnormal child psychology*, 29(6), 491–498. [PubMed: 11761283]
- Davis TL (1995). Gender differences in masking negative emotions: Ability or motivation? *Developmental Psychology*, 31(4), 660–667.
- Ebner NC (2008) Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods*. 40(1), 130–136. [PubMed: 18411535]
- Ebner NC, Riediger M, & Lindenberger U (2010) FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavioral Research Methods*, 42(1), 351–362.
- Eimer M, Holmes A, & McGlone FP (2003) The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, & Behavioral Neuroscience*, 3(2), 97–110.
- Ekman P, & Friesen WV (1982) Felt, false and miserable smiles. *Journal of nonverbal behavior*, 6(4), 238–252.

- Elfenbein HA (2006). Learning in emotion judgments: Training and the cross-cultural understanding of facial expressions. *Journal of Nonverbal Behavior*, 30(1), 21–36.
- Fotios S, Yang B, & Cheal C (2015). Effects of outdoor lighting on judgements of emotion and gaze direction. *Lighting Research & Technology*, 47(3), 301–315.
- Freyd JJ (1987) Dynamic mental representations. *Psychological Review*, 94(4), 427–238. [PubMed: 3317470]
- Gepner B, Deruelle C, & Grynfeldt S (2001) Motion and emotion: A novel approach to the study of face processing by young autistic children. *Journal of autism and developmental disorders*, 31(1), 37–45. [PubMed: 11439752]
- Gloor P (1997) *The temporal lobe and limbic system*. Oxford University Press, USA.
- Goeleven E, De Raedt R, Leyman L, & Verschuere B (2008) The Karolinska directed emotional faces: a validation study. *Cognition and Emotion*, 22(6), 1094–1118
- Golby AJ, Gabrieli JDE, Chiao JY, & Eberhardt JL (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience*, 4(8), 845–850. [PubMed: 11477432]
- Isaacowitz DM, Wadlinger HA, Goren D, & Wilson HR (2006). Is there an age-related positivity effect in visual attention? A comparison of two methodologies. *Emotion*, 6(3), 511–516. [PubMed: 16938091]
- Kamachi M, Bruce V, Mukaida S, Gyoba J, Yoshikawa S, & Akamatsu S (2013) Dynamic properties influence the perception of facial expressions. *Perception*, 42(11), 1266–1278. [PubMed: 24601038]
- Kunzmann U, Kappes C, & Wrosch C (2014). Emotional aging: a discrete emotions perspective. *Frontiers in psychology*, 5.
- Kunzmann U, Rohr M, Wieck C, Kappes C, & Wrosch C (2017). Speaking about feelings: Further evidence for multidirectional age differences in anger and sadness. *Psychology and aging*, 32(1), 93. [PubMed: 27977219]
- Krumhuber EG, Skora L, Kuster D, & Fou L (2017). A review of dynamic datasets for facial expression research. *Emotion Review*, 9(3), 280–292.
- LaBar KS, Crupain MJ, Voyvodic JT, & McCarthy G (2003) Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*, 13(10), 1023–1033. [PubMed: 12967919]
- Lamont AC, Stewart-Williams S, & Podd J (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition*, 33(6), 1017–1024. [PubMed: 16496722]
- Mather M, & Carstensen LL (2003). Aging and attentional biases for emotional faces. *Psychological Science*, 14(5), 409–415. [PubMed: 12930469]
- Mienaltowski A, Johnson ER, Wittman R, Wilson AT, Sturycz C, & Norman JF (2013). The visual discrimination of negative facial expressions by younger and older adults. *Vision research*, 81, 12–17. [PubMed: 23395863]
- Meissner CA, & Brigham JC (2001) Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3–35.
- Melinder A, Gredebäck G, Westerlund A, & Nelson CA (2010). Brain activation during upright and inverted encoding of own-and other-age faces: ERP evidence for an own-age bias. *Developmental science*, 13(4), 588–598. [PubMed: 20590723]
- Miner M, & Park DC (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4), 630–633.
- Noh SR, & Isaacowitz DM (2013). Emotional faces in context: age differences in recognition accuracy and scanning patterns. *Emotion*, 13(2), 238–249. [PubMed: 23163713]
- Orgeta V (2010). Effects of age and task difficulty on recognition of facial affect. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(3), 323–327.
- O’Toole AJ, Deffenbacher KA, Valentin D, & Abdi H (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, 22(2), 208–224. [PubMed: 8035697]
- O’Toole AJ, Roark DA, & Abdi H (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in cognitive sciences*, 6(6), 261–266. [PubMed: 12039608]

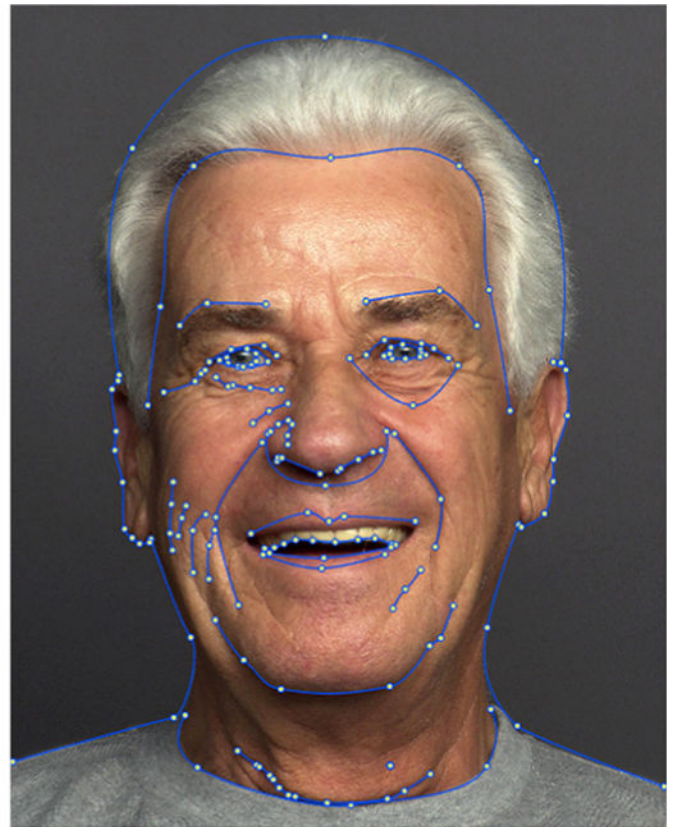
- Pascalis O, de Martin de Vivies X, Anzures G, Quinn PC, Slater AM, Tanaka JW, & Lee K (2011). Development of face processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 666–675. [PubMed: 22039564]
- Pascalis O, & de Schonon S (1995). Mother's face recognition by neonates: A replication and extension. *Infant Behavior and Development*, 18(1), 79–85
- Rhodes MG, & Anastasi JS (2012). The own-age bias in face recognition: a meta analytic and theoretical review. *Psychological bulletin*, 138(1), 146. [PubMed: 22061689]
- Riediger M, Voelkle MC, Ebner NC, & Lindenberger U (2011). Beyond "happy, angry, or sad?": Age-of-poser and age-of-rater effects on multi-dimensional emotion perception. *Cognition & emotion*, 25(6), 968–982. [PubMed: 21432636]
- Ruffman T, Henry JD, Livingstone V, & Phillips LH (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, 32, 863–881. [PubMed: 18276008]
- Samanez-Larkin GR, & Carstensen LL (2011) Socioemotional functioning and the aging brain In Decety J and Cacioppo J (Eds.) *The Oxford Handbook of Social Neuroscience* (pp. 507–521). New York: Oxford University Press.
- Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, ... & Nelson C (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research*, 168(3), 242–249. [PubMed: 19564050]
- Tran US, Lamplmayr E, Pintzinger NM, & Pfabigan DM (2013). Happy and angry faces: subclinical levels of anxiety are differentially related to attentional biases in men and women. *Journal of Research in Personality*, 47(4), 390–397.
- Voelkle MC, Ebner NC, Lindenberger U, & Riediger M (2012). Let me guess how old you are: effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging*, 27(2), 265. [PubMed: 21895379]
- Watier N, & Collin C (2012). The effects of distinctiveness on memory and metamemory for face-name associations. *Memory*, 20(1), 73–88. [PubMed: 22171812]
- Walker PM, & Tanaka JW (2003). An encoding advantage for own-race versus other-race faces. *Perception*, 32(9), 1117–1125. [PubMed: 14651324]



Figure 1.
Sample of face models from the FACES database selected to create the Dynamic FACES videos. Video versions for each sample available online.



Neutral



Happy

Figure 2.
Example of the MorphAge file setup used to create the Dynamic FACES videos

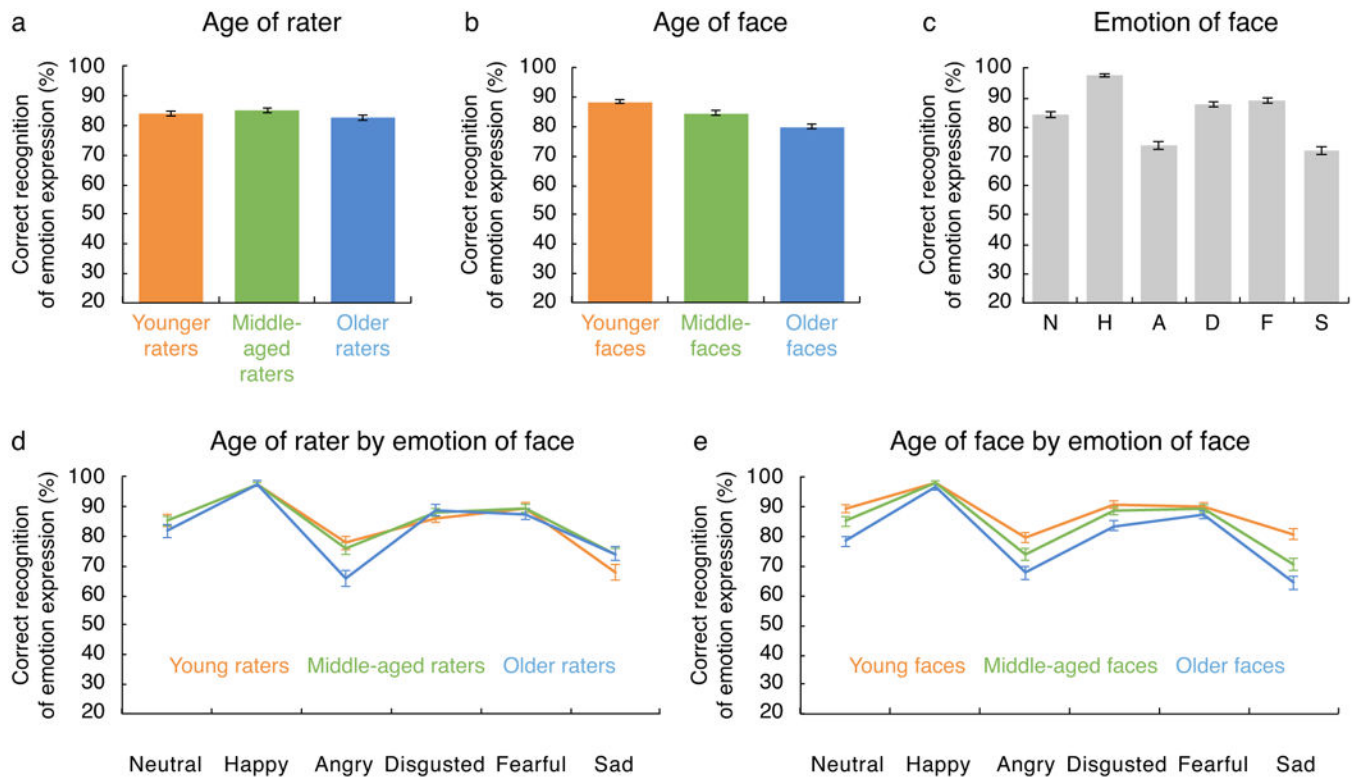


Figure 3. Effects on correct face emotion identification: (a) Main effect of rater age; (b) Main effect of face age; (c) Main effect of face emotion; N = neutral, H = happy, A = angry, D = disgusted, F = fearful, S = sad; (d) Interaction of rater age and face emotion; (e) Interaction of face age and face emotion. Error bars in all panels are 95% confidence intervals.

Table 1.

Face models included in the final Dynamic FACES database: distribution by age and gender

	Age 19–31	Age 39–55	Age 69–80
Women	29	27	29
Men	29	29	28

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Demographic characteristics of the sample of raters. SD = standard deviation

	Younger Raters (n = 639)		Middle-Aged Raters (n = 688)		Older Raters (n = 489)	
	Women	Men	Women	Men	Women	Men
Age Range	18–39		40–59		60–86	
Mean	29.8	29.9	50.8	50.3	68.5	68.9
SD	5.8	5.9	5.7	6.0	5.8	5.9
% 2 years of college or more	86.8	80.4	80.6	85.7	75.0	79.0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Estimated marginal means [and 95% confidence intervals] for emotion identification accuracy by rater age and face emotion.

	Neutral	Happy	Angry	Disgusted	Fearful	Sad	
Younger raters	85.2% [83.3%, 87.2%]	97.5% [96.7%, 98.3%]	77.8% [75.5%, 80.1%]	86.2% [84.5%, 88.0%]	89.5% [87.8%, 91.2%]	67.7% [65.2%, 70.2%]	84.1% [83.1%, 85.0%]
Middle-aged raters	85.2% [83.6%, 86.9%]	97.8% [97.1%, 98.5%]	76.0% [73.8%, 78.2%]	88.2% [86.7%, 89.7%]	89.4% [88%, 90.8%]	73.8% [71.6%, 76.0%]	85.1% [84.4%, 85.9%]
Older raters	81.8% [79.6%, 84.1%]	97.7% [96.7%, 98.7%]	65.5% [62.8%, 68.3%]	88.9% [87%, 90.8%]	87.6% [85.7%, 89.5%]	74.2% [71.8%, 76.6%]	82.6% [81.7%, 83.7%]
Older raters	84.3% [83.2%, 85.4%]	97.7% [97.2%, 98.2%]	73.8% [72.4%, 75.2%]	87.7% [86.7%, 88.7%]	88.9% [88.0%, 89.9%]	71.9% [70.5%, 73.2%]	

Table 4.

Estimated marginal means [and 95% confidence intervals] for emotion identification accuracy by face age and face emotion.

	Neutral	Happy	Angry	Disgusted	Fearful	Sad	
Younger faces	89.5% [88%, 91%]	98.3% [97.7%, 98.9%]	79.7% [77.8%, 81.6%]	91.0% [89.6%, 92.3%]	90.0% [88.6%, 91.4%]	80.9% [79.1%, 82.7%]	88.3% [87.6%, 88.9%]
Middle-aged faces	85.2% [83.5%, 86.9%]	98.2% [97.5%, 98.8%]	74.1% [72%, 76.1%]	88.8% [87.3%, 90.2%]	89.3% [87.9%, 90.7%]	70.7% [68.5%, 72.8%]	84.4% [83.7%, 85.2%]
Older faces	78.4% [76.5%, 80.3%]	96.6% [95.8%, 97.4%]	67.7% [65.5%, 69.8%]	83.6% [81.9%, 85.3%]	87.5% [86%, 89%]	64.3% [62.1%, 66.6%]	79.8% [79.0%, 80.6%]
	84.3% [83.2%, 85.4%]	97.7% [97.2%, 98.2%]	73.8% [72.4%, 75.2%]	87.7% [86.7%, 88.7%]	88.9% [88.0%, 89.9%]	71.9% [70.5%, 73.2%]	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript