Published in final edited form as:

Exp Aging Res. 2009 ; 35(2): 268–275. doi:10.1080/03610730902720638.

Lifespan Adult Faces: Norms for Age, Familiarity, Memorability, Mood, and Picture Quality

Kristen M. Kennedy, Kelly Hope, and Naftali Raz*

Department of Psychology and Institute of Gerontology, Wayne State University, Detroit, MI 48202.

Abstract

Because faces are frequently used as stimuli in cognitive aging behavioral and neuroimaging studies, there is a need for a normed set of pictures reflecting age and ethnic diversity. Minear and Park (2004) provided a large database of face pictures from average, ethnically diverse individuals across the adult lifespan. The current study evaluated 180 pictures from that database. 108 participants (mean age 23) rated each picture on specific attributes: perceived age, familiarity of appearance, mood, memorability, and image quality. Because the pictures did not differ in any of the measured attributes, we provide 3 balanced sets of face stimuli to researchers desiring normed face stimuli free from age confounds.

Keywords

aging; faces; standardized; norms; mood

Face stimuli are being increasingly used in studies of neuropsychology and in behavioral studies of memory, attention, perception, and emotion, and are frequently the stimuli of choice for use in functional neuroimaging studies (e.g., Farah, 1996; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Naveh-Benjamin, Guez, Kilb, & Reedy, 2004). Human faces are of particular interest in many regards. They are highly salient visual stimuli, with significant survival value in developmental (recognizing a caregiver) and social (friend or foe) contexts. Faces are also ecologically valid and familiar visual stimuli as all sighted humans are exposed to countless human faces in a lifetime. Furthermore, generating name-face memory associations is important in social transactions, and failure to retrieve a specific face in a face-name combination is a frequent complaint of elderly individuals (Riege, 1982). Older adults show significant impairment on tests involving episodic recall of faces (Grady et al., 1995; Crook & Larrabee, 1992). Moreover, better performance on face recall tests is observed when stimuli closely resemble the faces of the observers and persons with whom they interact (Golby, Gabrielli, Chiao, & Eberhardt, 2001). It was suggested that presentation of younger faces may give younger participants an advantage over their older counterparts (Bastin & Van der Linden, 2003), thus underscoring the importance of age in memory for faces. There is a distinct possibility that restricting experimental face stimuli to only younger faces introduces an unwanted confound to these studies.

A typical set of faces used in experimental studies are photographs of people between the ages of 18 and 50, and obtained from yearbook photos or convenient media sources. In such sets of stimuli, faces of older adults are scarce. The Minear & Park (2004) database was created to fill

*Correspondence concerning this article should be addressed to Naftali Raz, Ph.D., Institute of Gerontology, Wayne State University, 87 E Ferry, 226 Knapp Bldg, Detroit, MI 48202. Tel. (313) 577-2297; Fax: (508) 256-5689; electronic mail: nr@wayne.edu. Dr. Kennedy is now at the Center for BrainHealth, School of Behavioral and Brain Sciences at The University of Texas at Dallas.

this void. The database consists of a large representative sample of photographs of younger and older adults taken at public places accessible by all members of society (e.g. county fairs). The database contains 1,142 face-forward photographs of 576 people. In addition, right profiles are available for 308 individuals, and smiling pictures are available for 258 individuals. Of 576 participants, 281 are over the age of 50, and another 148 were older than 70 (Minear & Park, 2004). Furthermore, the sample represents diverse ethnic backgrounds.

However, this valuable ecologically valid database lacks the normative information that would assist in creating more balanced subsets as well as generating multiple parallel forms for use in retesting individuals. Therefore, the goal of the present study was to create such balanced subsets of stimuli by measuring specific characteristics of the stimuli as described below and equating the sets on those characteristics.

Method

Participants

The sample consisted of 108 undergraduate college students aged 18–54 years (mean age 22.58 ± 6.08 ; median age 21 years) from a large Midwestern urban university. The observers received course extra credit for their participation. The sample consisted of 88 women and 20 men, 51 were Caucasian, 36 African-American, and 21 listed their ethnicity as “other.” The participants’ ethnic background was equally distributed by age, $F(3,104) < 1, p = .64$. All participants provided their written informed consent and were debriefed in accord with university human investigation committee guidelines.

Procedure

The stimuli were sampled from the face database available at <http://agingmind.utdallas.edu/facedb> (Minear & Park, 2004). For this study, we chose 180 standardized black and white pictures for presentation on a computer terminal at 640×480 pixel resolution. In selecting these pictures, we used the following criteria: half of the faces were taken from the younger adults (ages 18–49) categories (neutral expression) and half from the older adults (ages 50–94) categories (neutral expression). In addition, half of the faces belonged to women and half belonged to men. We chose pictures that represented an ethnically diverse sample. The 180 pictures were divided into three sets of 60 pictures each equally balanced on age group, sex, and ethnic diversity.

All participants viewed all the pictures, with the order of presentation pseudo-randomized for each participant. Each image was displayed just above center of the screen and while the picture was on the screen, one question appeared below the picture. After the participant responded to the question, the next question appeared below the same picture, so the picture was available the entire time the participant was rating it. When all the questions for a specific picture were answered, the next picture was presented with the first question appearing below, and so on until all pictures had been seen and all questions answered.

Each picture was presented on the screen individually along with a question asking the participant to rate the picture on five specific attributes by using the Likert scale below the question with 1 indicating the least level of the attribute and 5 indicating the greatest level of that attribute. The five items were: Q1: How old do you think this person is? (*18–24, 25–30, 31–40, 41–49, 50–59, 60–69, 70–79, or 80 and older*); Q2: How familiar is this face? (*Not familiar at all, slightly unfamiliar, neutral, slightly familiar, or very familiar*); Q3: Does this person appear happy or sad? (*Sad, slightly sad, neutral, slightly happy, or happy*); Q4: How memorable is this face? (*Not at all memorable, slightly unmemorable, neutral, slightly memorable, very memorable*). Q5: Is this a good quality picture? (*Very poor quality, slightly*

poor quality, neutral, slightly good quality, or very good/excellent quality). Participants were instructed to use the mouse to indicate their choice (1, 2, 3, 4, or 5) for each question. The order in which the six questions were presented was randomized. Note that Q1 (rate the age of the face) was on an 8-pt scale.

The experiment was self-paced. When the participants responded to an item, they clicked on a button marked “Next” to invoke the next item. A screen appeared to notify the participants when they were 50% through the study and were allowed a short break if necessary.

Results

The 180 face stimuli were divided into three sets of 60 pictures each and Set served as a within-subject factor in the linear model. The dependent measure was the average Likert responses to each question (averaged across the 108 participants for each of the five attribute questions). A separate model was run for each attribute to determine if there were differences among the three sets of pictures. Descriptive statistics for each picture set are provided in the Table.

The mean Likert rating for the age of the person in the picture was 3.84 ($SD = 1.87$), which corresponds to approximately 45 years of age. The participants’ mean responses ranged from 1.18 which corresponds to the youngest age group (18–24 years) to 7.38 which falls in the next to oldest age group (70–79 years). The age rating did not differ among the three picture sets: $F(2, 177) < 1$.

The mean familiarity rating for the pictures was 2.83 ($SD = .25$), corresponding to “slightly unfamiliar” or rounded to “neutral”. The mean responses ranged from 2.24 to 3.69 across the pictures. Familiarity ratings did not differ among the three picture sets, $F(2, 177) < 1$. The mood of the persons in the pictures was rated on average 3.11 ($SD = .51$) or neutral and also did not differ by picture set, $F(2, 177) = 1.17$, *ns*. The mean mood ratings ranged from 2.00 to 4.68 across the pictures. The memorability of the pictures was rated 3.33 ($SD = .19$), or neutral, and did not differ among the picture sets, $F(2, 177) = 2.28$, $p = .11$. The mean memorability of the pictures ranged from 2.79 to 3.82. Finally, the mean picture quality ratings ranged from 2.46 to 3.84. The picture quality was rated on average as 3.00 ($SD = .22$), or exactly neutral, and ratings did not differ among the picture sets, $F(2, 177) < 1$.

Our ratings were also deemed to be fairly reliable. Across all questions (Q1–5): Cronbach’s $\alpha = .63$, Spearman-Brown reliability = .77 and split-half reliability $r = .62$. Because rated age (Q1) is a qualitatively, and quantitatively (8-point vs 5-point Likert scale) different item than the other four, we also report reliability for Q2–5: Cronbach’s $\alpha = .69$, Spearman-Brown = .80, and split-half reliability = .67. Among the attributes, coefficient of variation was largest (least rater agreement) for Q2 (familiarity) and smallest (most rater agreement) for Q1 (rated age).

To gauge the relation among the rated attributes we examined the correlations and found that the association strength varied across the attributes. Those correlations gauge the similarity among the attributes across the three sets (i.e., for all 180 pictures). As might be expected, the attributes of familiarity and memorability were significantly correlated, $r(178) = .54$, $p < .001$. Memorability and mood were significantly positively correlated, $r(178) = .48$, $p < .001$, as were mood and familiarity ($r(178) = .34$, $p < .001$), quality and familiarity ($r(178) = .47$, $p < .001$), quality and mood ($r(178) = .46$, $p < .001$), and quality and memorability ($r(178) = .43$, $p < .001$). Age did not correlate with mood, picture quality, or memorability (all p ’s $> .40$). Familiarity and age were significantly negatively correlated, $r(178) = -.20$, $p < .01$, with faces that were rated as older being rated as less familiar.

We then tested to see if there were differences in ratings according to characteristics of the raters (e.g., sex and ethnic background). There were a few differences: women rated the faces as slightly older ($F(1, 102) = 4.68, p < .05$), and more memorable ($F(1, 102) = 9.80, p < .01$) than men did. Familiarity also differed slightly by ethnicity of the raters: Asian participants rated faces higher in familiarity than raters who self-reported their ethnic background as multi-ethnic or other ($F(3, 102) = 3.30, p = .02$). No other associations were significant.

Discussion

We have created three balanced sets of face stimuli of average adult age, average familiarity, average picture quality, neutral mood, and average perceived memorability. We found that the 180 pictures (divided into three sets of 60 each) did not differ significantly on any of the five attributes: age of person in picture, familiarity of person in picture, mood of the person, picture quality, or memorability of the pictures. These findings justify using these ethnically diverse, age-varied, black and white pictures in computerized cognitive tests. Further, the association among the rated attributes was assessed. Rated age of the person in the picture was not associated with rated mood, picture quality, or memorability. Notably, stimuli with more positive rated affect were perceived as more memorable and of better quality. We also found that in this picture set, rated age was negatively associated with familiarity, indicating that faces that were rated as older were also rated as less familiar. This finding is likely due to the fact that we had a range of pictures in the norming set from early 20s to late 80s, whereas the participants were undergraduates who were on average 23 years old. Including older individuals in the norming sample may have provided different responses. A few demographic differences also existed with women rating faces as older on average and as more memorable than did men. There were also slight differences in rated familiarity by ethnic background. Overall, however, ratings of the faces across the five attributes, lists, and demographic groups of participants were remarkably similar.

Making sets of picture stimuli available, which have been normed on several attributes desirable to control for when creating cognitive tests, is a useful contribution to the field. It is becoming increasingly common, especially in studies of cognitive aging, to use faces as experimental stimuli (e.g., Naveh-Benjamin, Guez, Kilb, & Reedy, 2004). Numerous studies have investigated, for example, the effect of aging on recognition of unfamiliar faces (e.g., Bartlett, Leslie, Tubbs, & Fulton, 1989; Crook & Larrabee, 1992; Grady et al., 1995; Smith & Winograd, 1978). Neuroimaging studies using facial stimuli have found that visual processing of faces is “special” and is possibly processed in a different area of the brain (i.e., the “fusiform face area”; Golarai et al., 2007; Haxby et al., 1994; Ishai et al., 1999; Kanwisher, 2000; Kanwisher, McDermott, & Chun, 1997).

Being able to apply normed stimuli with neutral qualities to these types of studies will help ensure that any contextual background cues that may exist in some face stimuli do not exert an influence on the participant’s response. Further, use of the same standard set of picture stimuli among researchers simplifies comparisons across studies from different labs and across multiple cognitive domains (e.g., memory, attention, perception, priming).

Acknowledgments

This study was supported in part by a grant (R37 AG-11230) from National Institutes of Health. We thank Denise Park for the use of her database of face stimuli.

References

- Bartlett JC, Leslie JE, Tubbs A, Fulton A. Aging and memory for pictures of faces. *Psychology & Aging* 1989;4:276–283. [PubMed: 2803620]

- Bastin C, Van der Linden M. The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology* 2003;17:14–24. [PubMed: 12597069]
- Crook TH, Larrabee GJ. Changes in facial recognition memory across the adult life span. *Journal of Gerontology* 1992;47:138–14.
- Farah MJ. Is face recognition ‘special’? Evidence from neuropsychology. *Behavioral Brain Research* 1996;76:181–189.
- Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JDE, Grill-Spector K. Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience* 2007;10:512–522.
- Golby AJ, Gabrieli JDE, Chiao JY, Eberhardt JL. Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience* 2001;4:845–850.
- Grady CL, McIntosh AR, Horwitz B, Maisog JM, Underleider LG, Mentis MJ, Pietrini P, Schapiro MB, Haxby JV. Age related reductions in human recognition memory due to impaired encoding. *Science* 1995;269:218–221. [PubMed: 7618082]
- Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences* 1999;96:9379–9384.
- Kanwisher N. Domain specificity in face perception. *Nature Neuroscience* 2000;3:759–763.
- Kanwisher N, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 1997;17:4302–4311. [PubMed: 9151747]
- Minear M, Park DC. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers* 2004;36:630–633.
- Naveh-Benjamin M, Guez J, Kilb A, Reedy S. The associative memory deficit of older adults: further support using face-name associations. *Psychology & Aging* 2004;19:541–546. [PubMed: 15383004]
- Riege WH. Self-report and tests of memory aging. *Clinical Gerontologist* 1982;1:23–36.
- Smith AD, Winograd E. Adult age differences in remembering faces. *Developmental Psychology* 1978;14:443–444.
- Zacks, RT.; Hasher, L.; Li, KZH. *The Handbook of Aging & Cognition*. Vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates; 2000. Human memory; p. 293-357.

Table 1
Descriptive statistics for the three picture sets

	Median	Mean($\pm$ SE)	95% CI
List 1			
Age	3.896	3.947 ($\pm$.253)	3.441 – 4.453
Familiar	2.873	2.85 ($\pm$.032)	2.786 – 2.914
Mood	2.953	3.03 ($\pm$.064)	2.903 – 3.161
Quality	2.962	2.975 ($\pm$.025)	2.924 – 3.025
Memorable	3.368	3.368 ($\pm$.023)	3.322 – 3.415
List 2			
Age	3.311	3.711 ($\pm$.244)	3.223 – 4.200
Familiar	2.816	2.820 ($\pm$.028)	2.763 – 2.877
Mood	3.090	3.17 ($\pm$.071)	3.028 – 3.312
Quality	2.976	3.006 ($\pm$.029)	2.948 – 3.065
Memorable	3.307	3.332 ($\pm$.025)	3.281 – 3.382
List 3			
Age	3.660	3.877 ($\pm$.229)	3.419 – 4.335
Familiar	2.816	2.817 ($\pm$.037)	2.743 – 2.892
Mood	3.080	3.136 ($\pm$.063)	3.009 – 3.262
Quality	2.939	3.00 ($\pm$.030)	2.939 – 3.061
Memorable	3.311	3.293 ($\pm$.025)	3.242 – 3.342

Note. Age was rated on an eight point scale. Familiarity, Mood, Picture Quality, and Memorability were rated on a five point scale. SE – standard error of the mean; 95% CI – upper and lower confidence interval limits.