Journals of Gerontology: Psychological Sciences
cite as: *J Gerontol B Psychol Sci Soc Sci*, 2020, Vol. XX, No. XX, 1–10
doi:10.1093/geronb/gbaa038
Advance Access publication March 25, 2020

Research Article

Introducing a Face Sort Paradigm to Evaluate Age Differences in Emotion Perception

Katie Hoemann, MA,[1,*] Ishabel M. Vicaria, PhD,[1] Maria Gendron, PhD,[2] and Jennifer Tehan Stanley, PhD[3]

[1]Department of Psychology, Northeastern University, Boston, Massachusetts. [2]Department of Psychology, Yale University, New Haven, Connecticut. [3]Department of Psychology, The University of Akron, Ohio.

*Address correspondence to: Katie Hoemann, MA, Department of Psychology, Northeastern University, 360 Huntington Avenue, 125 Nightingale Hall, Boston, MA 02115-5005. E-mail: khoemann@gmail.com

Received: October 21, 2019; Editorial Decision Date: March 16, 2020

Decision Editor: Angela Gutchess, PhD

Abstract

Objectives: Previous research has uncovered age-related differences in emotion perception. To date, studies have relied heavily on forced-choice methods that stipulate possible responses. These constrained methods limit discovery of variation in emotion perception, which may be due to subtle differences in underlying concepts for emotion.
Method: We employed a face sort paradigm in which young ($N = 42$) and older adult ($N = 43$) participants were given 120 photographs portraying six target emotions (anger, disgust, fear, happiness, sadness, and neutral) and were instructed to create and label piles, such that individuals in each pile were feeling the same way.
Results: There were no age differences in number of piles created, nor in how well labels mapped onto the target emotion categories. However, older adults demonstrated lower consistency in sorting, such that fewer photographs in a given pile belonged to the same target emotion category. At the same time, older adults labeled piles using emotion words that were acquired later in development, and thus are considered more semantically complex.
Discussion: These findings partially support the hypothesis that older adults' concepts for emotions and emotional expressions are more complex than those of young adults, demonstrate the utility of incorporating less constrained experimental methods into the investigation of age-related differences in emotion perception, and are consistent with existing evidence of increased cognitive and emotional complexity in adulthood.

Keywords: Concepts and categories, Language, Research methods and issues, Social cognition

As people get older, they accrue considerable experience with building and maintaining social relationships. Accordingly, one might expect social perceptions, including emotion perception, to improve with age. However, age-related decline in performance on laboratory-based emotion perception tasks is widely observed, with meta-analyses (Gonçalves et al., 2018; Ruffman, Henry, Livingstone, & Phillips, 2008) reporting that older adults are significantly worse than young adults at identifying anger, fear, happiness, sadness, and surprise. These findings seem contradictory to the maintenance of healthy relationships (Lansford, Sherman, & Antonucci, 1998) in older age, which suggests intact social processing. In the present manuscript, we suggest that age-related differences in emotion perception may be due to subtle differences in conceptual knowledge about emotion, a crucial but understudied aspect of emotion perception that may improve with age (Labouvie- Vief, Grühn, & Studer, 2010; Magai, Consedine, Krivoshekova, Kudadjie-

Gyamfi, & McPherson, 2006). To address this gap, we introduce a face sort paradigm to evaluate age differences in emotion perception.

Testing Age Differences in Emotion Perception

To date, demonstrations of age differences in emotion perception have relied heavily on forced-choice methods that stipulate possible responses, and assume that deviations from expected responses represent deficits in emotion processing. In the typical task, perceivers view facial portrayals of emotion and select the "correct" emotion from a list of provided words. Researchers have noted that this task lacks ecological validity and have hypothesized that age differences may be attenuated by the use of more context-rich stimuli, which better approximate the real-world judgments older adults are familiar with and motivated to perform (Isaacowitz & Stanley, 2011). However, this hypothesis has not always been supported. Age differences in emotion perception are still found when using video clips compared to static images (Grainger, Henry, Phillips, Vanman, & Allen, 2015), implying that older adults' decline in performance is not due to a lack of dynamic information. Age differences are even found when participants are provided with training (Schlegel, Vicaria, Isaacowitz, & Hall, 2017), implying that older adults are not able to leverage task-relevant information to update (and improve) their judgments.

Another possibility is that the typical forced-choice format constrains responses. When participants are allowed to rate the extent to which a target is experiencing *multiple* emotions, older adults endorse more positive and neutral emotions for poses not intended to express those states (e.g., angry faces are rated as more neutral; Riediger, Voelkle, Ebner, & Lindenberger, 2011). They also attribute fewer negative emotions to intended portrayals (e.g., angry faces are rated as less angry; Riediger et al., 2011). Similarly, when participants are allowed to make multiple selections from a longer list of response options (17 words), older adults select more distinct words to describe neutral faces, compared to young adults (Kim, Geren, & Knight, 2015). These findings suggest that older adults see complexity in emotional targets that young adults do not.

One way of accounting for the observed age differences in emotion perception is to understand them as rooted in differences in the underlying concepts—a hypothesis consistent with a constructionist account of emotion. This account proposes that emotion perception is dependent upon conceptual knowledge, which is built through prior experience and shaped by language and culture (e.g., Barrett, Mesquita, & Gendron, 2011). In an instance of emotion perception, individuals bring emotion concepts—and the words that label them—to bear in inferring what another person is feeling. From this perspective, age group may serve as a type of developmental "culture" that molds the perception of emotion. Indeed, people hold different representations for how emotional faces should look at different ages (Fölster, Hess, & Werheid, 2014). More broadly, older adults may have more elaborated emotion knowledge because they have more accrued experience (Williams, Huang, & Bargh, 2009) and higher crystallized intelligence (Salthouse, 2009). These factors may provide older adults with concepts for emotional expressions that are more complex than those of young adults. Theoretical accounts of emotion across the life span propose that the ability to make subtle distinctions between emotions increases for healthy older adults (Labouvie- Vief et al., 2010; Magai et al., 2006); however, these accounts have focused on experience of emotion in the self, and have not been extended to the perception of emotion in others. Older adults have more elaborated vocabularies (Salthouse, 2009), and so may also have more semantically complex language for emotional expressions. Although previous research has found changes in the verbal expression of emotion across the life span (e.g., Magai et al., 2006), no study has examined the specific words being used by older versus young adults.

To test these hypotheses, however, an alternate emotion perception paradigm is necessary. Typical forced-choice emotion perception tasks often contain experimental features (such as a blocked trials design) and provide additional conceptual content (such as emotion words or scenarios) that participants can leverage to complete the task (e.g., Hoemann et al., 2019). In contrast, less constrained tasks can uncover nuances in emotion perception because they are not anchored on specific emotion categories. When participants are asked to freely label emotional stimuli, they provide a variety of responses beyond the expected emotion words (e.g., Betz, Hoemann, & Barrett, 2019). Overall performance on forced-choice tasks is significantly higher than on free-label alternatives, suggesting that the presence of emotion words in the task actively shapes emotion perception (Betz et al., 2019) and potentially obscures some between-group differences (Cassels & Birch, 2014). These differences may be especially relevant in an aging context because older adults are more likely than young adults to rely on heuristics provided by the environment when making decisions in order to reduce cognitive processing demands (Johnson, 1990; Mata, Schooler, & Rieskamp, 2007).

In particular, a face sort paradigm can be used to more directly test individuals' emotional inferences from faces. In this paradigm, participants are provided with a set of photographs of facial portrayals of emotion that belong to a set of target emotion categories (e.g., anger, sadness, happiness, etc.). Participants are asked to sort the photographs into piles, according to the emotion portrayed in each, and then label the piles with the corresponding emotion word(s) or phrase. A face sort paradigm provides the opportunity to discover subtle differences between older and young adults, as there may be differences in the number of piles created, the consistency of pile contents (i.e., how many photographs belong to the same target emotion category), and the number and nature of emotion words used as labels. This paradigm has previously been used, for example, to explore emotion perception in remote cultures (Gendron, Roberson, van der Vyver, & Barrett, 2014) and patients with the semantic variant of primary progressive aphasia (Lindquist, Gendron, Barrett, & Dickerson, 2014).

The Current Study

If, as we hypothesize, increased complexity in emotional inferences emerges across the life span due to accrued experience, then it is important to use an experimental paradigm that allows for perceiver-constructed rather than experimenter-stipulated emotion categories. Given that the majority of studies on emotion perception and aging have used a forced-choice paradigm, it is likewise important to investigate the extent to which findings hold up when obtained with a less constrained response format. Even if the overall pattern of age differences is similar, an alternate paradigm is useful for triangulating the construct of emotion perception. Indeed, if age differences replicate across different ways of measuring emotion perception, researchers can have more confidence that their findings reflect true age differences, rather than an artifact or influence of the response format.

In the current study, we compared young and older adults' sorting of a target set of facial portrayals of emotion in a face sort paradigm. These data were previously collected as part of a larger study described in Stanley and Isaacowitz (2015); the current analyses were not preregistered. While the older adult data were used as a control sample by Lindquist and colleagues (2014), the young adult data have not previously been reported, and young and older adults have not previously been compared. Based on previous literature, we made two predictions. First, we predicted that older adults would evidence finer-grained sorting performance than young adults (as suggested by Kim et al., 2015; Riediger et al., 2011), as demonstrated by a larger number of piles formed, the consistency of which may also be affected. Second, we predicted that older adults would evidence more elaborated vocabularies for emotional portrayals than young adults (as suggested by Magai et al., 2006; Salthouse, 2009), as measured in terms of how well pile labels mapped onto the target emotion categories, and in terms of their semantic complexity.

Method

Participants

Participants were 42 young adults (YA) and 43 older adults (OA); see Table 1 for demographics. Fifty-two YA and 57 OA were originally recruited for the parent study (Stanley & Isaacowitz, 2015). Ten YA and 14 OA were excluded from analysis due to missing or incomplete data or lack of fluency in English. One OA was excluded for creating 23 piles (the only outlier). There were no age differences based on gender ($X^2(1) = 0.30$, $p > .05$, $\varphi = 0.06$), racial category ($X^2(3) = 3.62$, $p > .05$, $\varphi = 0.21$), or self-reported health ($t(70) = 1.13$, $p > .05$, $d = 0.27$). OA had more years of education, $t(83) = 3.53$, $p < .01$, $d = 0.77$, and demonstrated higher vocabulary scores on the 20-item Shipley Vocabulary Test (Zachary, 1986), $t(82) = 2.63$, $p < .05$, $d = 0.57$ (replicating previous results; Salthouse, 2009).

A power analysis in G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) confirmed that the total sample size of 85 provided adequate power .80 (max value 1) to detect medium ($f = 0.25$) within- and between-main effects, and within–between interactions, for two age groups and six emotion category measurements. This estimated effect size is consistent with prior studies using forced-choice tasks (e.g., Ruffman et al., 2008).

Stimuli

Photographs of 120 facial portrayals of emotion (20 White, young adult posers; 10 male, 10 female) were sourced from the Karolinska (Lundqvist, Flykt, & Öhman, 1998) and IASLab Face Set databases. Faces portrayed six target emotion categories (anger, disgust, fear, happiness, sadness, and

Table 1. Participant Demographics

	N	Age (years)	Gender	Racial category	Years of education	Self-reported health (1 to 5)	Vocabulary score
Young adults	42	21.71 (3.37)	45% male 55% female	83% White 12% Asian 2% Native American 2% Other	14.95 (2.30)*	4.02 (0.81)	13.88 (1.98)*
Older adults	43	73.91 (5.85)	51% male 49% female	95% White 5% Asian	16.86 (2.67)*	3.80 (0.85)	15.09 (2.24)*

Note: Number of participants in each group represent those submitted for analysis in the present study. Means are given for continuous variables, with standard deviations in parentheses.
$^*p < .05$.

neutral) and were printed in full color on 4 × 6 inch laminated cards.

Procedure

Participants were provided with a stack of the 120 photographs in a randomized order and were asked to sort them into piles based on feeling, such that individuals who were feeling the same way were placed together. The task was untimed and participants were free to modify their piles. The verbatim instructions to participants were:

> In this pile there are pictures of a bunch of people. All of the people in the pictures are feeling something. The people in the pile are feeling different things. What I would like you to do is to sort the people in this pile based on how they feel. You can create as many piles as you need to. At the end, each pile should have people who all feel the same way. Please go as slowly as you need. This will not be timed. Feel free to examine the images before you begin sorting.

At the end of the sorting period, participants were asked to provide the experimenter with a label for each pile to describe the feeling that was shared by those people. The experimenter recorded the pile labels and, after the participant left, which photographs belonged to each pile.

Analysis

Descriptive analysis of sorting performance

To gather descriptive information on how young and older adults sorted the face stimuli, we first conducted separate cluster analyses on the data from each age group (as outlined in Gendron et al., 2014). To prepare the data for the cluster analysis, we computed a co-occurrence matrix for each group (Coxon & Davies, 1982). Each co-occurrence matrix contained a row and column for each of the 120 items in the set, resulting in a 120 × 120 symmetrical matrix. Each cell in the matrix represented the number of times a given pair of face items was grouped by participants in the same pile (i.e., across participants). The larger the number in a cell, the more frequently those two items co-occurred, and thus the higher perceived similarity between those items at a group level. We then converted this co-occurrence similarity matrix into a distance matrix, in which a higher cell value was an indication of less similarity between items. The cluster analysis was then performed on each dissimilarity matrix (one for YA, one for OA).

We chose a hierarchical cluster analysis (Sokal & Michener, 1958) in which the number of clusters is discovered rather than prespecified (unlike other clustering procedures such as *k*-means; Hartigan & Wong, 1979). We used an agglomerative approach, starting with each item as its own cluster and progressively linking items together based on estimates of their distance from one another (computed from the number of times face stimuli appeared in the same vs separate piles). We employed an average linkage clustering method because it uses information about all pairs of distances to assign cluster membership, not just the nearest or farthest item pairs (Aldenderfer & Blashfield, 1984). To determine an appropriate number of clusters for each group's solution, we examined the resulting dendrograms as well as the clustering coefficients at each stage of agglomeration.

Categorizing facial portrayals

Next, we directly compared the sorting performance of young and older adults in terms of the number of piles created, and the consistency of those piles. To calculate the consistency scores, we first determined which emotion comprised the greatest number of faces in each pile, based on the posers' intended portrayals (e.g., anger, fear, disgust, happiness, sadness, or neutral). We then calculated a consistency score to reflect the percentage of faces matching the dominant emotion of each pile.

Labeling facial portrayals

Lastly, we directly compared the labeling performance of young and older adults in terms of how well pile labels mapped onto the target emotion categories, and in terms of their semantic complexity. Pile labels were transformed into adjectival form (i.e., "anger" became "angry") and stripped of any modifiers (e.g., "very"). Two trained coders classified labels by emotion category. Labels containing multiple words within the same target category (e.g., "fearful and scared") were counted as belonging to that category. Labels that contained one word belonging to a target category and one nontarget word (e.g., "scared and unsure") or words from two different target categories (e.g., "scared and disgusted") were counted as not belonging to the target set. We observed "surprise" labels to be common in both age groups, so we counted them separately from other nontarget responses. Coding reliability was acceptable (Cohen's kappa = 0.83), and discrepancies were resolved through discussion before analysis.

To examine potential age differences in semantic complexity, we gathered three measures for each word used as a pile label: age of acquisition (AoA; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), number of meanings (i.e., senses; Princeton University, 2010), and log frequency of use (Davies, 2008–). These measures have previously been used as indices of semantic complexity and to model semantic growth (e.g., Steyvers & Tenenbaum, 2005), and are reported in Supplementary Tables S1 and S2. Values were averaged for piles labels containing more than one word, yielding one measure of AoA, number of meanings, and frequency per pile.

Results

Descriptive Analysis of Sorting Performance

Separate agglomerative hierarchical cluster analyses for each age group revealed that the largest increase in the clustering

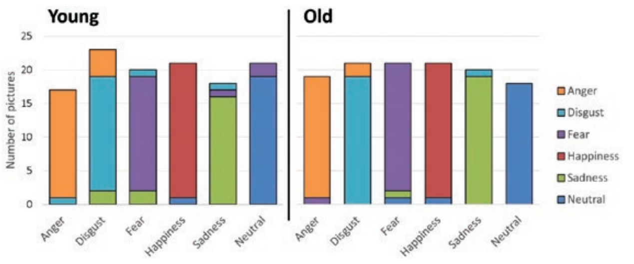

Figure 1. Six-cluster solutions derived from a hierarchical agglomerative cluster analysis for young adult (left panel) and older adult (right panel) participants' free sorting data. Clusters are plotted on the *x*-axis, and the *y*-axis represents the number of pictures grouped into a given cluster. Clusters are labeled based on the predominant emotion and presented alphabetically. Stacked bars (i.e., bars containing several different shades) indicate that pictures portraying different emotions were clustered together.

coefficient (i.e., the within-cluster average item distance) occurred between five and six clusters, indicating that the six-cluster solutions divided the items more coherently. These six-cluster solutions revealed that both age groups' cluster solutions converged on the six target emotions. The six-cluster solution for each age group are illustrated in Figure 1. As can be seen, YA and OA sorting data yielded nearly identical contents for each of the identified clusters. However, we also observed that young adults' data converged to the six-cluster solution with more internal consistency (i.e., less idiosyncratic variation) than did the older adults' data, as demonstrated graphically by shorter within-cluster bracket lengths in the dendrogram (Supplementary Figure S1).

Categorizing Facial Portrayals

Number of piles

There were no age differences in the number of piles created, $t(83) = -0.24$, $p > .05$, $d = 0.06$. Although there were only six target emotion categories in the stimulus set, on average YA created 7.33 piles ($SD = 1.30$; minimum = 6, maximum = 12, range = 6) and OA created 7.40 piles ($SD = 1.90$; minimum = 4, maximum = 15, range = 11).

Consistency

A mixed-design analysis of variance, with consistency as the dependent variable, emotion as the repeated measure (anger, disgust, fear, happiness, sadness, and neutral), and age group (YA vs OA) as the between-participants factor revealed a significant main effect of age group, $F(1,83) = 23.00$, $p < .001$, $\eta_p^2 = 0.22$, such that OA had lower consistency in sorting emotion faces than did YA. There was also a significant main effect of emotion, $F(5,415) = 46.18$, $p < .001$, $\eta_p^2 = 0.36$ (Huynh-Feldt corrected at $e = 0.91$), which was qualified by an age by emotion interaction, $F(5,415) = 3.00$, $p < .05$, $\eta_p^2 = 0.04$. Post hoc *t* tests revealed that OA consistency was significantly lower than YA for all but one emotion ($ps < .05$)—effect sizes (d) ranging from 0.54 for sadness to 0.93 for fear—the exception being disgust, $t(83) = 1.78$,

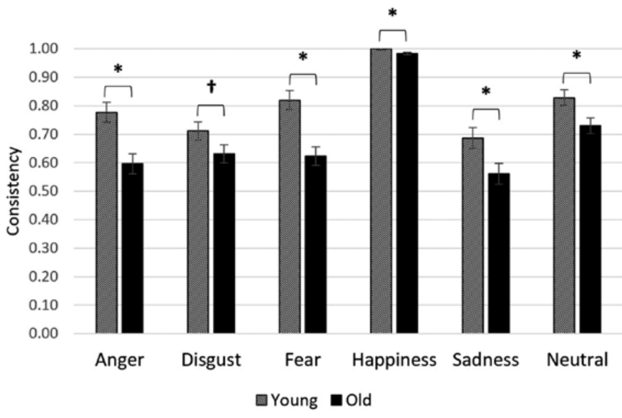

Figure 2. Follow-up comparisons between young adult (YA; dashed lines) and older adult (OA; solid bars) sorting consistency. Consistency was measured by first determining the dominant emotion (anger, disgust, fear, happiness, sadness, or neutral) of each pile as that which comprised the greatest number of faces, and then calculating the percentage of faces matching the dominant emotion of each pile. $*p < .05$; $^{\dagger}p < .10$ (Bonferroni adjusted).

$p > .05$, $d = 0.38$ (all *p*s Bonferroni adjusted; see Figure 2). The distribution of cross-category sorts (errors) by emotion for each age group is provided in Table 2, where the total errors is the inverse of consistency (i.e., 1 − consistency) for a given emotion category. The patterns of errors are similar to forced-choice tasks (Ruffman et al., 2008), showing anger is incorrectly categorized as disgust (and vice versa) by both young and older adults. This confusion can also be seen in Figure 1, with the disgust cluster for both age groups containing a portion of angry faces.

Labeling Facial Portrayals

Mapping to target emotion categories

A chi-square test indicated that the distribution of pile labels per emotion category (Table 3) did not differ by age group, $X^2(8) = 6.77$, $p > .05$, $\varphi = 0.10$, such that there were no differences between how well YA or OA pile labels mapped onto the six targets.

Semantic complexity

We analyzed differences in the semantic complexity of YA and OA labels using hierarchical linear modeling (HLM; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004), with piles (Level 1) nested within individuals (Level 2). We ran a series of models to examine the effect of age group (Level-2 predictor) on label AoA, number of meanings, and frequency (Level-1 outcome variables; see Supplementary Material for details). Age group was a significant predictor of AoA ($b = 0.35$, $SE = 0.17$, $t(82) = 2.04$, $p < .05$, $R^2 = 0.09$[1]), such

[1] Estimate of the proportional reduction in variance, or percent of variance accounted for by the predictor, calculated as ($\mathrm{Var}[u_0\mathrm{NoPredictor}] - \mathrm{Var}[u_0\mathrm{Predictor}]/\mathrm{Var}[u_0\mathrm{NoPredictor}]$) following Peugh (2010).

Table 2. Distribution of Errors Per Emotion Category

		Anger piles	Disgust piles	Fear piles	Happiness piles	Sadness piles	Neutral piles	Other piles	Total errors
Young	Anger faces	—	12%	3%	0%	5%	1%	2%	23%
	Disgust faces	15%	—	5%	0%	0%	1%	8%	29%
	Fear faces	1%	6%	—	0%	3%	1%	7%	18%
	Happiness faces	0%	0%	0%	—	0%	0%	0%	0%
	Sadness faces	4%	8%	15%	0%	—	3%	1%	31%
	Neutral faces	1%	0%	9%	5%	4%	—	2%	17%
Old	Anger faces	—	16%	4%	0%	4%	5%	11%	40%
	Disgust faces	17%	—	4%	0%	1%	4%	10%	37%
	Fear faces	7%	8%	—	3%	6%	6%	8%	38%
	Happiness faces	0%	1%	0%	—	1%	0%	0%	2%
	Sadness faces	10%	10%	11%	0%	—	6%	6%	44%
	Neutral faces	4%	1%	4%	10%	3%	—	5%	27%

Note: Piles (columns) are categorized according to the dominant emotion, or the emotion represented by the majority of faces in a given pile (note that if a participant had more than one pile in which a given emotion dominated, these piles have been combined for the present analysis). The "Other" column represents piles for which there was not a single dominant emotion. Photograph stimuli (rows) were categorized according to the emotion portrayed. Cell values represent the mean percentage of errors for each pile–emotion combination (i.e., of the 20 photographs for each portrayed emotion, how many were sorted into piles corresponding to a different dominant emotion).

Table 3. Percentage of Participants Who Produced Pile Labels Per Emotion Category

	Target emotion categories						Nontarget responses		
	Anger	Disgust	Fear	Happiness	Sadness	Neutral	Surprise	Other	Combination
Young	88%	88%	57%	98%	88%	90%	74%	40%	23%
Old	81%	70%	44%	98%	88%	77%	86%	51%	30%

that OA produced labels that were acquired later (M = 6.33, SD = 0.71) than YA did (M = 6.06, SD = 0.56). For example, OA produced "anxious," "sorrowful," and "vexed," versus YA "scared," "sad," and "mad" (word clouds for both age groups are presented in Supplementary Figures S2 and S3). Neither the relationship between age group and frequency (b = −0.15, SE = 0.09, t(82) = −1.66, p > .05, R^2 = 0.05), nor that between age and number of meanings (b = −0.06, SE = 0.10, t(82) = −0.63, p > .05, R^2 = 0.02) was significant.

To control for known age differences in vocabulary (Salthouse, 2009), we tested a model for AoA including a Level-2 predictor for standardized vocabulary score as well as an age by vocabulary interaction term. Neither vocabulary (b = 0.05, SE = 0.09, t(80) = 0.54) nor the interaction term (b = −0.02, SE = 0.06, t(80) = −0.36) was a significant predictor of AoA (both ps > .05, combined incremental R^2 = 0.02). Although OA demonstrated significantly higher vocabulary scores than YA in our sample, that difference does not appear to be driving label generation.

To explore the relationship between semantic complexity and the identification of intended facial portrayals, we examined the bivariate correlation between participants' mean AoA and sorting consistency. We found that individuals who produced labels that were acquired later in development also evidenced lower sorting performance, r = −0.21, p = .05.

Discussion

In the present study, young and older adults were asked to sort 120 photographs of faces according to the emotion portrayed in each. Contrary to our prediction that older adults would make finer-grained distinctions, both age groups generated a similar number of piles, and labeled these piles using words that mapped similarly onto the six target emotion categories (anger, disgust, fear, happiness, sadness, and neutral). However, differences emerged in two important ways. First, older adults demonstrated significantly lower sorting performance in terms of the consistency of pile contents per emotion category. Second, consistent with our prediction that older adults would make use of more a more elaborated vocabulary for emotional expressions, older adults labeled piles using emotion words that were more semantically complex (i.e., were acquired later in development). Further, older adults' use of semantically complex words was significantly related to differences in sorting performance. These findings provide partial support for our hypothesis that older adults' concepts for emotions and emotional expressions are more complex than those of young adults, and demonstrate the utility of incorporating less constrained experimental methods into the investigation of age-related differences in emotion perception.

Our findings extend the existing literature on emotion perception in several ways. By comparing the consistency of pile contents between young and older adults, we observed a common pattern of age-related differences: while older adults evidenced lower sorting performance overall, this effect was not significant for disgust faces. Inconsistencies in age-related performance on disgust are well-documented, with meta-analyses finding little to no difference between young and older adults (Gonçalves et al., 2018; Ruffman et al., 2008). The present study replicates these findings in a new experimental paradigm, and offers insights into the types of labels that participants freely generated for piles containing disgust images. Overall, there was little lexical variation: "disgusted" was used in all but two instances, the exceptions being "grossed out" (used by a young adult) and "revolted" (used by an older adult). These observations add depth by suggesting that disgust varies little across the life span in terms of specificity and complexity.

The subtle age differences seen in this task were revealed by using a more open-ended, participant-driven approach to emotion perception. This approach may be particularly important when studying age differences, because older adults are more likely than young adults to rely on cues in the environment, thereby reducing the amount of cognitive processing required for a judgment (Johnson, 1990; Mata et al., 2007). This could mean that older adults are more constrained by forced-choice tasks than young adults. As this area of research becomes more advanced, moving beyond accuracy criteria toward investigating the conceptual understanding of emotions can uncover important distinctions and similarities. It may seem like a step backward to measure emotion perception using a more exploratory and time-intensive approach. Our findings, however, suggest that it is worth the investment. While consistency of face sorting yields a performance measure similar to that obtained in forced-choice tasks, it is enriched by other aspects of the data, such as labeling and group-level clustering, that allow for novel insights and a more complete comparison of young versus older adults. Where applicable, both face sort and forced-choice methods may be used to triangulate results.

Limitations and Considerations

Subtle differences between older and young adults' emotion perception may further be revealed by studies that make use of more nuanced, naturalistic stimuli. One limitation of the present study is that the stimuli used were all photographs of young adults, so the observed age differences in sorting consistency may be partially due to how older adults perceive out-group emotions. At the time the data collection for Stanley and Isaacowitz (2015) was designed, in 2009, validated databases of older adults' facial portrayals of emotion were not yet available (e.g., the FACES database; Ebner, Riediger, & Lindenberger, 2010). Additionally, the present study used only photographs of White posers, limiting our conclusions about the role of ethnicity. Although the set of photographs drew from two separate face databases, and were both presented in color, there may be features of these stimuli (e.g., background, hair, clothing) that influenced participants' choices (cf. Silver & Bilker, 2015). Specific features of the present face sort task may have also constrained performance. For example, each poser appeared in six photographs (one for every target emotion), which may have tacitly encouraged participants to sort stimuli so that the same individual never appeared in the same pile twice. To determine if the present findings generalize, it is important for future studies to take these sorts of experimental features into account.

Likewise, the conclusions of the present study may be limited by the fact that our young and older adult participants were not matched for vocabulary and years of education. Although age-related differences in vocabulary are well-established (Salthouse, 2009), it could be that this contributed to the enhanced semantic complexity of older adults' pile labels. However, our analyses showed no such effect. Similarly, although it is not uncommon for older adult samples to have more years of education than their young adult counterparts (e.g., Noh & Isaacowitz, 2013; Silver & Bilker, 2015), this may have also affected performance. Previous studies have found that participants with college degrees outperform participants without college degrees on forced-choice emotion perception tasks (Trauffer, Widen, & Russell, 2013), and that larger age-related differences in the perception of disgust are associated with larger differences in years of education (Gonçalves et al., 2018). In the present study, the older adults had significantly more years of education, and so may have had an advantage. Yet our results suggest otherwise: older adults still evidenced lower sorting performance, including on disgust (although the difference did not reach significance). Moreover, both age groups averaged between 14 and 17 years of education, indicating at least some college education.

Implications for Future Research

Age differences in emotion perception are routinely interpreted as deficits due to positivity effects (Carstensen, Fung, & Charles, 2003), general cognitive decline (Salthouse, 2009), or reductions in the structure or function of certain brain regions (Gonçalves et al., 2018; Ruffman et al., 2008), such that older adults have more difficulty identifying intended emotions. An alternate interpretation is that older adults' accrued experience may render them more likely to perceive complexity in stimuli (Kim et al., 2015). From this perspective, the present findings are consistent with a constructionist account of emotion perception, in which different experiences and language abilities (e.g., vocabulary) are not epiphenomenal to emotion perception but rather actively contribute to individual and group differences (e.g., Barrett et al., 2011). A young adult without the concept "vexed," for example, may be more likely to access "angry" for a facial portrayal intended as anger. An older adult with a more developed conceptual repertoire may first access "vexed" for that same facial portrayal, and the precision of this label may correspond with different sorting performance.

The link between language and perception proposed by a constructionist account of emotion is based on evidence that perceivers implicitly activate concepts (and their associated labels) during tasks, such that language provides an automatic and inherent context for emotion perception (e.g., Gendron, Lindquist, Barsalou, & Barrett, 2012; Nook, Lindquist, & Zaki, 2015). Indeed, in the present study we observed that older adults' use of semantically complex pile labels was tied to lower consistency in pile contents, suggesting that the complexity of older adults' emotion concepts may drive sorting performance. Nevertheless, without direct evidence of causal influence on perception, it is possible that the language differences we observed in the present study do not reflect differences in underlying emotion concepts. In particular, it is possible that these findings were due to a cohort effect: the current generation of older adults may have always used words like "vexed," whereas the current generation of young adults has not. If this were the case, then age group may not be a developmental "culture" that molds individual experience, after all. Instead, differences in emotion concepts may be due to differences in each cohort's formative historical and sociopolitical contexts. To test this possibility, future research could construct equivalent corpora of language from different eras (e.g., newspaper reports, magazine articles) to investigate if the words used to describe emotion have changed over time.

The interpretation that older adults perceive more complexity in emotional stimuli is also consistent with existing evidence of increased cognitive and emotional complexity across the life span (Grühn, Lumley, Diehl, & Labouvie-Vief, 2013, but see Charles, Piazza, & Urban, 2017), and thereby with accounts such as Differential Emotions Theory (Magai et al., 2006) and Dynamic Integration Theory (Labouvie- Vief et al., 2010) which predict this increase as a function of development in knowledge about mental states and how they manifest in the self and others. In particular, older adults have been shown in some studies to have higher emotional granularity than young adults (e.g., Carstensen et al., 2011; Ready, Carvalho, & Weinberger, 2008), suggesting that they represent emotional experience with greater nuance and specificity via a more elaborated conceptual system for emotion (but see Hay & Diehl, 2011; Ong & Bergeman, 2004). These findings are bolstered by recent evidence that older adults differentiate more among the meanings of negative emotion words (Ready, Santorelli, & Mather, 2019). It is an open question whether emotional granularity, and emotional complexity more broadly, influences emotion perception (e.g., Brooks & Freeman, 2018). Future research should continue to examine the interrelationship between complexity and age, and consider these factors within the context of emotion perception.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* online.

Funding

This work was supported by the National Institute on Aging (grant numbers T32 AG-00204, R01AG-026323); and the National Heart, Lung, and Blood Institute (grant number 1 F31 HL140943-01 to K. Hoemann).

Acknowledgments

We are grateful to Hanna Negami, Amy Stricoff, and Nia Fogelman for assistance with data collection. All data analyzed and reported in this paper are available via a data repository hosted by the Center for Open Science at https://osf.io/ea3dm/. Development of the Interdisciplinary Affective Science Laboratory (IASLab) Face Set was supported by the National Institutes of Health Director's Pioneer Award (DP1OD003312) to Lisa Feldman Barrett. More information is available online at www.affective-science.org.

Conflict of Interest

None reported.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Newbury Park, CA: Sage.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, **20**, 286–290. doi:10.1177/0963721411422522

Betz, N., Hoemann, K., & Barrett, L. F. (2019). Words are a context for mental inference. *Emotion*, **19**, 1463–1477. doi:10.1037/emo0000510

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour,* **2**, 581–591. doi:10.1038/s41562-018-0376-6

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. doi:10.1023/A:102456980

Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., . . . Nesselroade, J. R. (2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging*, **26**, 21–33. doi:10.1037/a0021285

Cassels, T. G., & Birch, S. A. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading' task: Implications for measuring perspective-taking and emotion recognition. *PLoS One*, **9**, e93653. doi:10.1371/journal.pone.0093653

Charles, S. T., Piazza, J. R., & Urban, E. J. (2017). Mixed emotions across adulthood: When, where, and why? *Current Opinion in Behavioral Sciences,* **15**, 58–61. doi:10.1016/j.cobeha.2017.05.007

Coxon, A. P. M., & Davies, P. (1982). *The user's guide to multidimensional scaling: With special reference to the MDS(X) library of computer programs*. Exeter, NH: Heinemann Educational Books.

Davies, M. (2008–). The corpus of contemporary American English (COCA): 560 million words, 1990-present. Retrieved from https://corpus.byu.edu/coca/

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods,* **42**, 351–362. doi:10.3758/BRM.42.1.351

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods,* **41**, 1149–1160. doi:10.3758/BRM.41.4.1149

Fölster, M., Hess, U., & Werheid, K. (2014). Facial age affects emotional expression decoding. *Frontiers in Psychology,* **5**, 30. doi:10.3389/fpsyg.2014.00030

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion,* **12**, 314–325. doi:10.1037/a0026007

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion,* **14**, 251–262. doi:10.1037/a0036052

Gonçalves, A. R., Fernandes, C., Pasion, R., Ferreira-Santos, F., Barbosa, F., & Marques-Teixeira, J. (2018). Effects of age on the identification of emotions in facial expressions: A meta-analysis. *PeerJ,* **6**, e5278. doi:10.7717/peerj.5278

Grainger, S. A., Henry, J. D., Phillips, L. H., Vanman, E. J., & Allen, R. (2015). Age deficits in facial affect recognition: The influence of dynamic cues. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, **72**, 622–632. doi:10.1093/geronb/gbv100

Grühn, D., Lumley, M. A., Diehl, M., & Labouvie-Vief, G. (2013). Time-based indicators of emotional complexity: Interrelations and correlates. *Emotion,* **13**, 226–237. doi:10.1037/a0030363

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **28**, 100–108. doi:10.2307/2346830

Hay, E. L., & Diehl, M. (2011). Emotional complexity and emotion regulation across adulthood. *European Journal of Aging*, **8**, 157–168. doi:10.1007/s10433-011-0191-7

Hoemann, K., Crittenden, A. N., Msafiri, S., Liu, Q., Li, C., Roberson, D., . . . Feldman Barrett, L. (2019). Context facilitates performance on a classic cross-cultural emotion perception task. *Emotion,* **19**, 1292–1313. doi:10.1037/emo0000501

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

Johnson, M. M. (1990). Age differences in decision making: A process methodology for examining strategic information processing. *Journal of Gerontology,* **45**, P75–P78. doi:10.1093/geronj/45.2.p75

Kim, S., Geren, J. L., & Knight, B. G. (2015). Age differences in the complexity of emotion perception. *Experimental Aging Research,* **41**, 556–571. doi:10.1080/0361073X.2015.1085727

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods,* **44**, 978–990. doi:10.3758/s13428-012-0210-4

Labouvie-Vief, G., Grühn, D., & Studer, J. (2010). Dynamic integration of emotion and cognition: Equilibrium regulation in development and aging. In M. E. Lamb, A. M. Freund, & R. M. Lerner (Eds.), *The handbook of life-span development, vol. 2: Social and emotional development* (pp. 79–115). Hoboken, NJ: John Wiley & Sons, Inc.

Lansford, J. E., Sherman, A. M., & Antonucci, T. C. (1998). Satisfaction with social networks: An examination of socioemotional selectivity theory across cohorts. *Psychology and Aging,* **13**, 544–552. doi:10.1037//0882-7974.13.4.544

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion,* **14**, 375–387. doi:10.1037/a0035293

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces (KDEF)*. Stockholm, Sweden: Department of Clinical Neurosciences, Karolinska Hospital.

Magai, C., Consedine, N. S., Krivoshekova, Y. S., Kudadjie-Gyamfi, E., & McPherson, R. (2006). Emotion experience and expression across the adult life span: Insights from a multimodal assessment study. *Psychology and Aging,* **21**, 303–317. doi:10.1037/0882-7974.21.2.303

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging,* **22**, 796–810. doi:10.1037/0882-7974.22.4.796

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

Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion,* **15**, 569–578. doi:10.1037/a0039166

Ong, A. D., & Bergeman, C. S. (2004). The complexity of emotions in later life. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences,* **59**, 117–122. doi:10.1093/geronb/59.3.p117

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology,* **48**, 85–112. doi:10.1016/j.jsp.2009.09.002

Princeton University. (2010). WordNet. Princeton University. Retrieved from https://wordnet.princeton.edu/

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: SSI, Inc.

Ready, R. E., Carvalho, J. O., & Weinberger, M. I. (2008). Emotional complexity in younger, midlife, and older adults. *Psychology and Aging,* **23**, 928–933. doi:10.1037/a0014003

Ready, R. E., Santorelli, G. D., & Mather, M. A. (2019). Older and younger adults differently judge the similarity between negative affect terms. *Aging & Mental Health,* **23**, 325–328. doi:10.1080/13607863.2017.1421614

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

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

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, **30**, 507–514. doi:10.1016/j.neurobiolaging.2008.09.023

Schlegel, K., Vicaria, I. M., Isaacowitz, D. M., & Hall, J. A. (2017). Effectiveness of a short audiovisual emotion recognition training program in adults. *Motivation and Emotion*, **41**, 646–660. doi:10.1007/s11031-017-9631-9

Silver, H., & Bilker, W. B. (2015). Colour influences perception of facial emotions but this effect is impaired in healthy ageing and schizophrenia. *Cognitive Neuropsychiatry*, **20**, 438–455. doi:10.1080/13546805.2015.1080157

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, **38**, 1409–1438.

Stanley, J. T., & Isaacowitz, D. M. (2015). Caring more and knowing more reduces age-related differences in emotion perception. *Psychology and Aging*, **30**, 383–395. doi:10.1037/pag0000028

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, **29**, 41–78. doi:10.1207/s15516709cog2901_3

Trauffer, N. M., Widen, S. C., & Russell, J. A. (2013). Education and the attribution of emotion to facial expressions. *Psychological Topics*, **22**, 237–247.

Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology*, **39**, 1257–1267. doi:10.1002/ejsp.665

Zachary, R. A. (1986). *Shipley Institute of living scale revised manual*. Los Angeles, CA: Western Psychological Services.