Vision Research 157 (2019) 97–104

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Classification images characterize age-related deficits in face discrimination

Sarah E. Creighton[a], Patrick J. Bennett[a,*], Allison B. Sekuler[b,c,a]

[a] Department of Psychology, Neuroscience, and Behaviour, McMaster University, Hamilton, Ontario L8S 4K1, Canada
[b] Rotman Research Institute, Baycrest Health Sciences, Toronto, Ontario M6A 2E1, Canada
[c] Department of Psychology, University of Toronto, Toronto, Ontario M5S 3G3, Canada

ARTICLE INFO

Number of reviewers - 1

Keywords:
Face perception
Aging
Classification image
Efficiency
Noise

ABSTRACT

Face perception is impaired in older adults, but the cause of this decline is not well understood. We examined this issue by measuring Classification Images (CIs) in a face discrimination task in younger and older adults. Faces were presented in static, white visual noise, and face contrast was varied with a staircase to maintain an accuracy rate of ≈71%. The noise fields were used to construct a CI using the method described by Nagai et al. (2013) and each observer's CI was cross-correlated with the visual template of a linear ideal discriminator to obtain an estimate of the absolute efficiency of visual processing. Face discrimination thresholds were lower in younger than older adults. Like Sekuler, Gaspar, Gold, and Bennett (2004), we found that CIs from younger adults contained structure near the eyes and brows, suggesting that those observers consistently relied on information conveyed by pixels in those regions of the stimulus. CIs obtained from older adults were noticeably different: CIs from only two older adults exhibited structure near the eye/brow regions, and CIs from the remaining older observers showed no obvious structure. Nevertheless, face discrimination thresholds in both groups were strongly and similarly correlated with the cross-correlation between the CI and the ideal template, suggesting that despite older observers' lack of consistent structure, the CI method is sensitive to between-subject differences in older observers' perceptual strategy.

1. Introduction

Older adults are impaired on a number of face perception tasks, such as eyewitness identification (Searcy, Bartlett, & Memon, 1999), detecting manipulations to facial features (Konar, Bennett, & Sekuler, 2013; Murray, Halberstadt, & Ruffman, 2010; Slessor, Riby, & Finnerty, 2012), identifying faces varying in expression (Ruffman, Henry, Livingstone, & Phillips, 2008), and/or viewpoint (Habak, Wilkinson, & Wilson, 2008; Lee, Grady, Habak, Wilson, & Moscovitch, 2011), and interpreting social cues (Halberstadt, Ruffman, Murray, Taumoepeau, & Ryan, 2011), yet the cause of these age-related declines is poorly understood. One possibility is that age differences in face perception are caused by older and younger adults relying on different sources of information, or different kinds of processing, to identify faces. For example, younger adults (Goffaux & Dakin, 2010; Pachai, Sekuler, & Bennett, 2013b) rely more heavily on horizontal than vertical structure to identify faces. But the reliance on horizontal structure is diminished with older observers (Goffaux, Poncin, & Schiltz, 2015; Obermeyer, Kolling, Schaich, & Knopf, 2012; Sekuler, Pachai, Creighton, & Bennett, 2014; Yu & Chung, 2011). The critical horizontal structure is concentrated around the eye/brow region (Duncan et al., 2017; Pachai, Sekuler, & Bennett, 2013a), a region that is particularly important for face identification (Gold, Sekuler, & Bennett, 2004; Gosselin & Schyns, 2001; Pachai et al., 2013a; Schyns, Bonnar, & Gosselin, 2002; Sekuler, Gaspar, Gold, & Bennett, 2004) and some emotions (Duncan et al., 2017; Smith, Cottrell, Gosselin, & Schyns, 2005). Thus, age-related deficits in face processing may be due to the decreased efficiency with which older observers encode task-relevant information conveyed by the eye/brow region. This idea is consistent with behavioural studies showing older adults are less sensitive at discriminating changes in the spacing between the eyes than are younger adults (Chaby, Narme, & George, 2011; Slessor et al., 2012).

Further support for an age-related deficit in processing information around the eyes is provided by several eye-tracking studies. Despite poorer recognition memory, older adults make more eye movement transitions between inner facial features (such as the eyes) than do younger adults (Chan, Kamino, Binns, & Ryan, 2011; Firestone, Turk-Browne, & Ryan, 2007). Similarly, as faces become better learned, younger adults make fewer fixations to faces, and spend more time on critical regions (e.g., eyes) and relatively less time on uninformative regions (e.g., nose, mouth); however, older adults show no change in their distribution of fixations, and spend more time on uninformative

* Corresponding author.
E-mail addresses: creighs@mcmaster.ca (S.E. Creighton), bennett@mcmaster.ca (P.J. Bennett), asekuler@research.baycrest.org (A.B. Sekuler).

https://doi.org/10.1016/j.visres.2018.07.002
Received 26 October 2017; Received in revised form 16 July 2018; Accepted 16 July 2018
Available online 01 August 2018
0042-6989/

Vision Research 157 (2019) 97–104

regions such as the nose (Heisz & Ryan, 2011). Finally, older adults spend proportionately less time looking toward the upper- than lower-halves of upright angry, fearful, and sad faces compared to younger adults (Murphy & Isaacowitz, 2010; Wong, Cronin-Golomb, & Neargarder, 2005), and this difference in gaze pattern is correlated with identification accuracy on at least some of these emotions (Murphy & Isaacowitz, 2010; Sullivan, Ruffman, & Hutton, 2007; Wong et al., 2005). However, these age differences in eye movements should be interpreted with caution because where one fixates may not provide a complete picture of the information observers use to perform a perceptual task.

Another potential explanation is that age-related declines in face processing are related to age differences in holistic processing, the extent to which faces are encoded and represented as a unified, integrated whole, rather than a simple linear summation of its component parts (see Maurer, Le Grand, & Mondloch, 2002 for a review suggesting the importance of holistic processing for face perception; but see Gaspar, Bennett, & Sekuler, 2008; Konar, Bennett, & Sekuler, 2010; Sekuler et al., 2004 for an alternative view). Support for this hypothesis is equivocal: the results of some studies are consistent with impaired holistic processing in older adults (Chaby et al., 2011; Murray et al., 2010; Schwarzer, Kretzer, Wimmer, & Jovanovic, 2010; Slessor et al., 2012), whereas other results suggest that holistic processing does not decline (Boutet & Faubert, 2006) or perhaps becomes even more important for identifying faces and other objects, even as face discrimination performance declines (Adduri & Marotta, 2009; Creighton, Sekuler, & Bennett, 2012; Daniel & Bentin, 2012; Dror, Schmitz-Williams, & Smith, 2005; Konar et al., 2013; Pilz, Bennett, & Sekuler, 2010).

One potential explanation for these equivocal findings is the presence of large individual differences or heterogeneity within the older adult population. Such heterogeneity would suggest that a full understanding of age-related differences in face processing would require examining how face encoding strategies differ across individuals. In the current paper, we begin to address this issue using response classification to derive classification images (CIs) of individual observers. Specifically, here we use the random sub-sampling variant of the CI method developed by Nagai et al. (2013), which allows for the estimation of stable classification images in far fewer trials than the traditional method (≈1500 vs ≈10,000; Gold et al., 2004; Sekuler et al., 2004), making it an ideal approach for studying special populations, such as older adults. To our knowledge, only one published paper (Éthier-Majcher, Joubert, & Gosselin, 2013) has used a related technique to explore judgements of facial trustworthiness in healthy older adults.

The response classification paradigm has several advantages over methods used in the studies described above: (1) it allows us to examine individual differences in face processing strategies in populations where substantial within-group variability might obscure between-group effects; (2) it does not require long duration stimuli; (3) it makes no *a priori* assumptions about regions of interest; thus it can reveal unexpected strategies not captured by other methods (e.g., feature displacement, composite faces, eye-tracking); and (4) it potentially reveals the relationship between an observer's response and perceptual strategy, to the extent that a given observer demonstrates a consistent relationship, while holding performance approximately constant across individuals. This last point is particularly relevant in cases where we expect significant differences to exist across individuals and groups.

In a typical response classification experiment, on each trial a unique external noise field is generated and added to one of two randomly chosen signals (e.g., Face A or Face B). The observer then must determine which stimulus was shown. For signals embedded in a high level of external noise, on some trials, by chance, the external noise will be distributed in such a way that Face A more closely resembles Face B, making the observer more likely to misclassify the stimulus as Face B. On other trials, the external noise may amplify characteristics of Face A that lead to a more likely Face A classification. After many trials, the contrast values of the noise fields are sorted into a 2 × 2 stimulus-response matrix, and the fields within each bin are averaged and combined to form a classification image (CI) using the equation:

$$CI = (\overline{N_{AA}} + \overline{N_{BA}}) - (\overline{N_{AB}} + \overline{N_{BB}})$$

where $\overline{N_{SR}}$ denotes the pixelwise average of the noise fields for all trials in a given stimulus-response class (e.g., $\overline{N_{AB}}$ represents the average of all noise fields for trials when the stimulus was Face A and the observer responded Face B). For a linear observer, the expected value of the CI is proportional to the observer's template (Murray, Bennett, & Sekuler, 2005). In other words, classification images can be conceptualized as "behavioural receptive fields" that provide a visual representation of how each region of the stimulus influences an individual observer's response (Gold et al., 2004). Using this approach, we can determine if older and younger adults use qualitatively similar sampling strategies (i.e., rely on the same information when discriminating faces), at the level of the individual observer, rather than at the level of age-groups; and examine the extent to which age differences in classification images are associated with age differences in face discrimination thresholds.

2. Methods

2.1. Observers

Ten younger ($M = 23.3$, $SD = 3.56$) and ten older ($M = 71.7$, $SD = 6.95$) Caucasian adults participated in this experiment either for cash ($10/hour) or partial course credit. Older adults were screened for visual pathologies with a vision and general health questionnaire. The Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) and the Montreal Cognitive Assessment (Nasreddine et al., 2005) were used to screen for cognitive impairments in older adults; all MMSE and MoCA scores fell within the normal range. All participants had normal or corrected-to-normal Snellen visual acuity, and the groups also did not differ significantly on peak contrast sensitivity, as measured by the Pelli-Robson chart (Table 1). All but two participants (SEC and MVP) were unpracticed psychophysical observers with no prior exposure to the stimuli, and were naïve to the purpose of the experiment.

2.2. Apparatus

Stimuli were generated on an Apple Macintosh G5 PowerPC (OS 10.5.8) using MATLAB (v 7.4.0) and the Psychophysics and Video toolboxes (Brainard, 1997; Pelli, 1997). Stimuli were presented on a NEC monitor (36 cm × 27 cm) with a resolution of 640 × 480 pixels and frame rate of 85 Hz. The average luminance of the display was 67 cd/m^2. The display was the only source of illumination in the testing room. Participants viewed the stimuli binocularly at a distance of 88 cm while seated in an adjustable chair. A chin rest was used to stabilize head position throughout the experiment.

2.3. Stimuli

The stimuli were two Caucasian male faces from the Gold, Bennett,

Table 1
Mean (SD) age, near and far Snellen decimal acuity, Pelli-Robson contrast sensitivity, Mini-Mental State Exam (MMSE), and Montreal Cognitive Assessment (MoCA).

	N	Age	Near Acuity	Far Acuity	Pelli-Robson	MMSE	MoCA
Younger	10	23.3 (3.56)	1.3 (0.29)	1.3 (0.26)	1.97 (0.05)		
Older	10	71.7 (6.95)	0.9 (0.17)	1.0 (0.12)	1.88 (0.13)	28.9 (1.52)	26.9 (2.28)

and Sekuler (1999) face set. These stimuli were chosen in order to be consistent with Nagai et al. (2013), who first used the random sub-sampling method in a special population. The stimuli were equated for amplitude spectrum and differed only in their phase spectrum (see Gold et al., 1999, for more details about the stimuli). Each face was centred in a 128×128 pixel array ($4.41^\circ \times 4.41^\circ$), and the height and width of each face subtended a visual angle of $3.41^\circ \times 2.41^\circ$, respectively.

We used the random sub-sampling method first described by Nagai et al. (2013), which has been shown to yield similar face classification images as fully-sampled faces, but in a fraction of the trials. To create the sub-sampled stimuli, all pixels falling outside an elliptical mask surrounding the face were set to zero contrast. One pixel from each 2×2 pixel region of the face was randomly selected for presentation; the remaining three pixels in that region were set to zero contrast. The spatial location of the presented pixels was held constant within an observer and across sessions, but varied across observers. On each trial, a unique noise field was generated by randomly selecting contrast values from a Gaussian distribution with a mean of zero and a standard deviation of 0.3. Noise values more than two SD from the mean were resampled until all values were within range. The final image was created by adding this noise field to the face stimulus. Note that the noise fields contained non-zero values only at pixels that were shown in the sub-sampled faces. Fig. 1 illustrates this stimulus generation process. In the experiment, contrast of the face stimulus was adjusted according to a single 2-down/1-up staircase maintaining response accuracy of approximately 71%.

2.4. Procedure

The McMaster University Research Ethics Board approved the experimental protocol, and written informed consent was obtained from all subjects prior to their participation.

The procedure was verbally explained to each participant, and task instructions were presented on the screen. Participants adapted to the average luminance of the display by fixating the centre of the computer screen for 60 s, and then completed two blocks of 20 practice trials to familiarize them with the task. In the first practice block, root mean square (RMS) face contrast was set to 0.3, and did not vary; in the second block, contrast was adjusted with a 2-down/1-up staircase maintaining response accuracy of approximately 71%.

Each trial began with a central fixation dot that was displayed for 1000 ms, followed by an individual face + noise stimulus for 506 ms, a blank screen for 200 ms, and then a response screen consisting of the two possible faces displayed at high contrast, without sub-sampling. Response screen stimuli were the same dimensions as target stimuli, centred 2.21° to the left and right of the centre of the display. Participants performed a face discrimination task by pressing one of two keys on a computer keyboard to indicate whether the stimulus was Face A (presented on the left) or Face B (presented on the right). Auditory feedback indicated whether the response was correct (high tone) or incorrect (low tone), and then the fixation dot appeared to indicate the beginning of the next trial. Unlimited response time was given, and participants were aware that the probability of either face appearing on a given trial was 50%.

Each session consisted of 1500 experimental trials, and participants were allowed a self-timed break every 100 trials. The two sessions usually were completed on consecutive days, with each session lasting approximately 90 min.

3. Results

As in Nagai et al. (2013), analyses were based only on subsampled pixels (i.e., pixels displayed to the participant). Classification images were estimated by sorting the contrast values of the noise fields presented on each trial into a stimulus-response matrix. The first 50 trials of each session were treated as practice trials and discarded, and classification images were estimated from the remaining 1450 trials per session for a total of 2900 trials. Thresholds were calculated as the average of the last 200 staircase reversals in each session. Analyses were performed in MATLAB (v 7.4.0) and R (R Development Core Team, 2016).

Further analyses quantified which pixels were correlated with observers' decisions. First, to increase the signal:noise ratio, smoothed classification images were created by convolving the compressed raw classification images with a 10×10 uniform convolution kernel. One consequence of this process is that pixel contrast is no longer independent, therefore calculating statistical significance levels using standard statistical tests is inappropriate. We therefore used the following permutation test to evaluate the null hypothesis that the values in the smoothed classification image could be produced by chance. The responses of each observer were shuffled randomly, and a series of 10 new, smoothed classification images were computed with the shuffled responses. Because there was no association between stimulus noise and the shuffled response, values in the smoothed classification images were used to estimate the distribution of CI pixel values that arise when the null hypothesis is true (Efron & Tibshirani, 1993). Statistical significance was determined by comparing the observed CI pixel values in an observer's filtered classification image with this null distribution. The Bonferroni correction for multiple comparisons tends to be too conservative, particularly when there are many tests that are not independent (Bland & Altman, 1995). Therefore, we chose instead to set the threshold for significance to $p < .001$ to facilitate comparisons with previous studies that used similar methods and stimuli.

Fig. 2 shows the spatial location of significant pixels (red regions, $p < .001$) in the smoothed classification images of each observer in each age group for each individual session and combined across both sessions. Additionally, Fig. 3 shows the z-scored version of younger (top) and older (bottom) observers' CIs for the combined sessions. By the end of session 2 (Fig. 2), CIs from younger adults exhibit significant pixels mainly in the eye/brow regions, the primary exception being the CI from observer JNW, which had few significant pixels. By contrast, even after 2900 trials, the CIs from most older adults had few significant pixels, and the small number of significant pixels were distributed more broadly across the face and to less informative regions, such as the forehead (GEZ, 54% of all significant pixels), nose (HUG, 60%), or mouth (MOF and MRF, 43% and 22% respectively). The CIs from two older observers (BAG & GEZ) did exhibit structure near the eye/brow regions (61% & 42%), but they also exhibited structure at less informative regions such as the cheek (BAG, 29%) and forehead (GEZ, 54%).

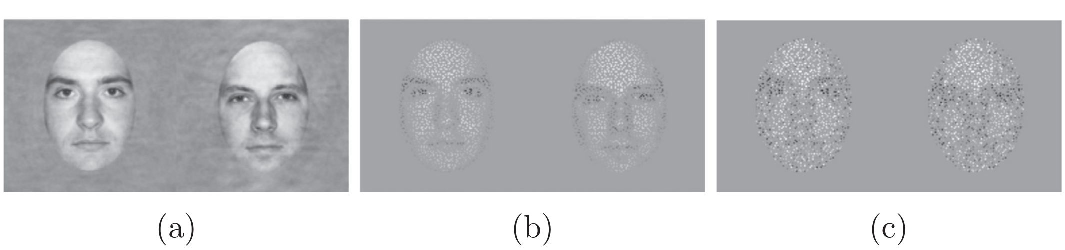


Fig. 1. Stimulus generation: (a) fully-sampled, (b) sub-sampled, (c) sub-sampled plus noise.

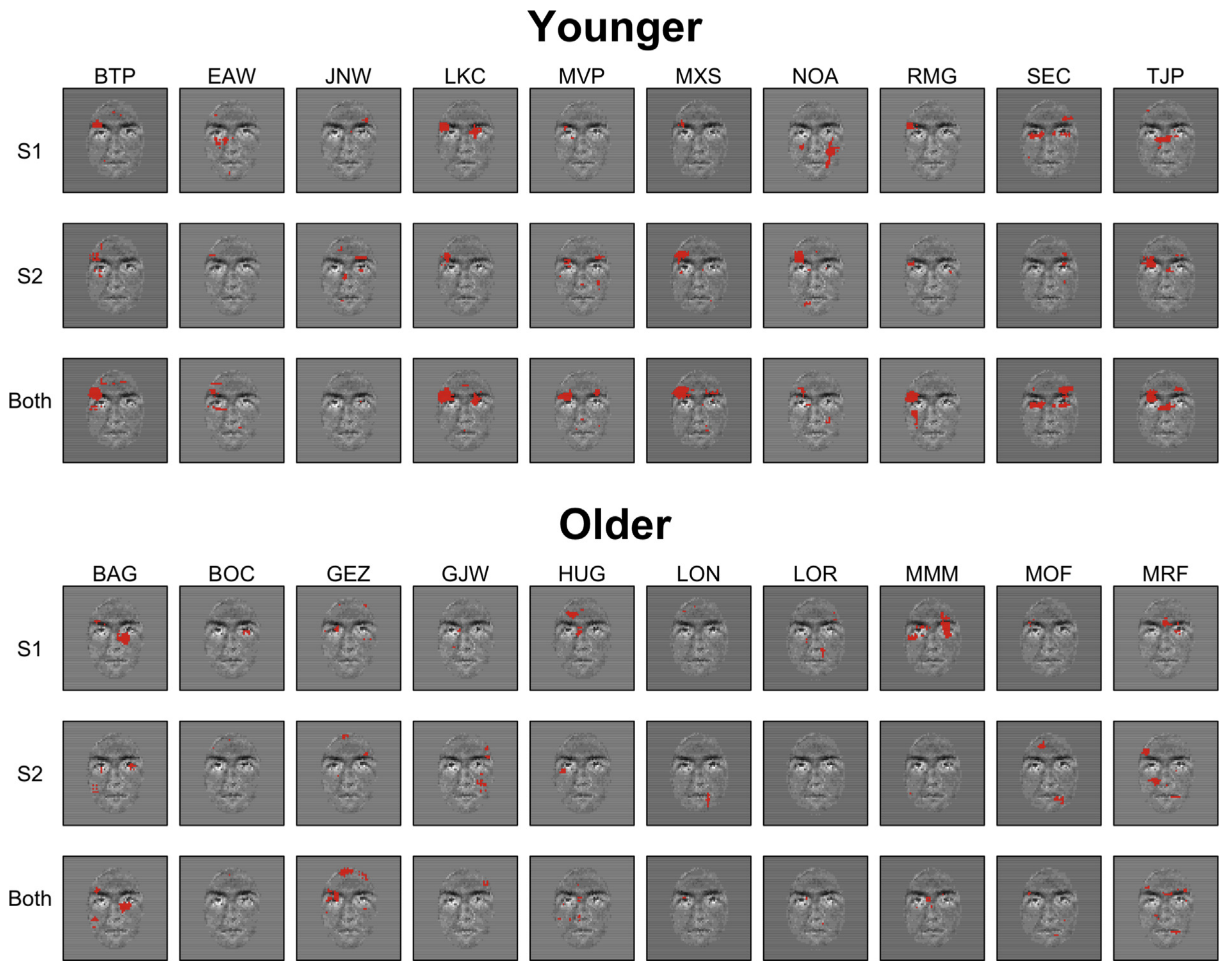


Fig. 2. Significant pixels ($p < .001$) in smoothed classification images for younger (top panel) and older (bottom panel) observers for each session alone (S1, S2) and after 2900 trials (both).

Fig. 5a and b show the spatial distribution and density of significant pixels at the group level for the combined sessions. Note that these maps are only approximate in spatial location (a consequence of the sub-sampling being random for each subject) with the presented pixel in each 2×2 region treated as coming from the same position. Younger observers' decisions were nearly exclusively driven by the eye/brow region (Fig. 5a), and showed considerable within-group consistency, with 7 of 10 observers relying on the left eye/brow. On the other hand, older observers demonstrated greater spatial variability and less consistency (Fig. 5b), with at most 2 of 10 observers relying on the same pixel location. Fig. 5c illustrates the overlap in classification images across age groups, by highlighting only pixels that were statistically significant in at at least one younger and one older subject. This analysis suggests that, at the group level, classification images from younger and older observers did overlap to some degree, with observers in both groups relying on information conveyed by pixels near the eyes and brows. However, this overlap was driven nearly entirely by just two older observers, GEZ and BAG.

In younger adults, the distribution of significant pixels was relatively stable across sessions 1 and 2 (Fig. 2). This stability is consistent with the results of Nagai et al. (2013), who found that as few as 1450 trials were sufficient to obtain stable classification images in younger adults using the sub-sampling method employed here. On the other hand, older adults demonstrated less consistency across sessions. For example, in session 1 observer HUG used pixels in the forehead and nose region, but in session 2 HUG used a region near one eye; in session 1, the significant pixels for observer MOF were located in the left eye/brow region, but in session 2 they were located on the forehead and

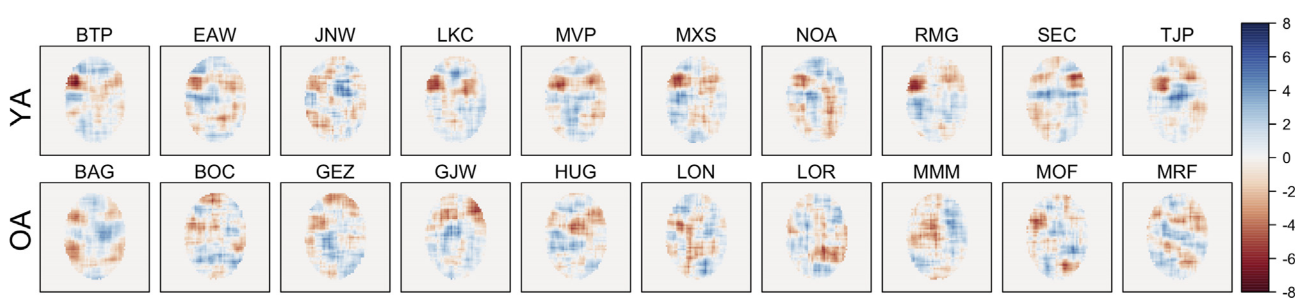


Fig. 3. Z-scored smoothed classification images for younger (YA, top panel) and older (OA, bottom panel) adults after 2900 trials. CIs were z-scored based on the individual distribution obtained from each observers' permutation test.

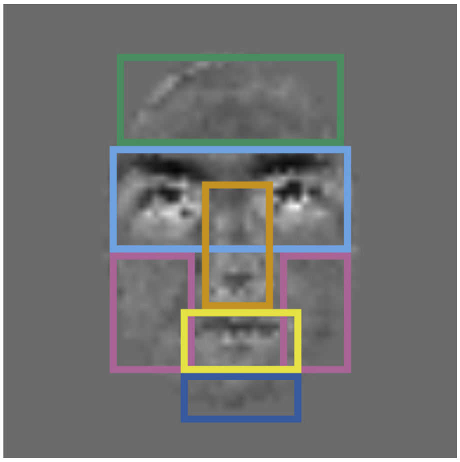


Fig. 4. Areas defined for the region of interest analysis. Anatomical features include the eyes, nose, mouth, forehead, cheeks, and chin. Note that significant pixels in the left and right cheek were summed for the analysis. Line width has been increased in the figure for the sake of visibility.

mouth; and observer MMM had significant pixels in the right and left eye/brow regions in session 1, but only one significant pixel in those regions in session 2, with a few additional significant pixels between the eyebrows by the end of 2900 trials.

Table 2 lists face discrimination threshold (RMS contrast) and the number of significant pixels in the smoothed classification image for each participant. Thresholds, averaged across sessions, were significantly higher in older observers than younger observers ($t(10.17) = 3.85$, $p = .002$, one-tailed). Additionally, consistent with our observations from Fig. 2, classification images for older adults contained fewer significant pixels compared to younger adults, even after 2900 trials ($t(16.87) = 3.21$, $p = .003$, one-tailed). We performed a region of interest analysis by separately counting the number of significant pixels within the eyes, nose, mouth, forehead, cheeks, and chin (see Fig. 4 for anatomically defined regions). For both groups, the majority of significant pixels were contained in the eye region ($M_Y = 0.82$, $M_O = 0.52$) relative to other facial features (nose: $M_Y = 0.05$, $M_O = 0.11$; mouth: $M_Y = 0.01$, $M_O = 0.05$; forehead: $M_Y = 0.20$, $M_O = 0.27$; cheeks: $M_Y = 0.06$, $M_O = 0.13$; chin: $M_Y = 0.00$, $M_O = 0.00$). As expected, younger adults relied more heavily on the eye region than did older adults ($t(12) = 3.94$, $p = .001$, one-tailed). No age difference was found for the nose, mouth, forehead, cheeks, or chin (all $ps > .21$). Pairwise t tests indicated that younger adults had more significant pixels in the eyes compared to other regions (all $ps < .003$). Older observers tended to rely more on the eyes than the mouth and chin ($ps < .04$).

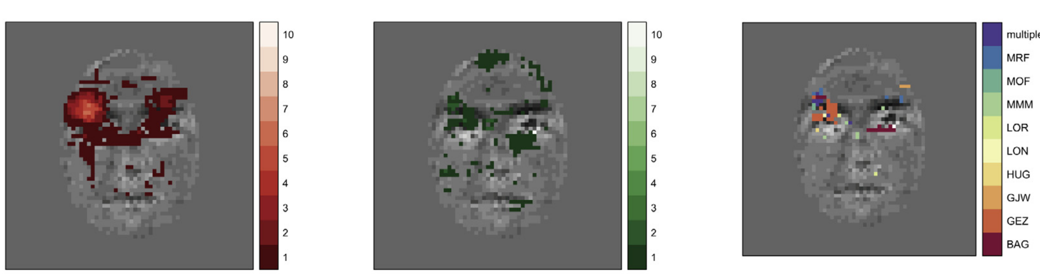
Table 2 also lists the cross-correlation of the observer's CI with the ideal template. If the classification image captures the important aspects of an observer's perceptual strategy, then we would expect absolute efficiency – as indexed by the cross-correlation between the CI and the ideal linear template – to be associated with an observer's threshold (Murray et al., 2005). Threshold is plotted as a function of the cross-correlation between the CI and ideal template in Fig. 6. Threshold and cross-correlation values are significantly correlated overall ($r = -0.89$, $p < .001$, one-tailed) and in both younger ($r = -0.74$, $p = .007$, one-tailed) and older ($r = -0.73$, $p = .009$, one-tailed) participants. Additionally, averaged across sessions, older observers had significantly lower cross-correlations relative to younger observers ($t(14.48) = 4.75$, $p < .001$, one-tailed). We also estimated absolute efficiency for each observer by computing the squared ratio of human to ideal d' when face RMS contrast was set to the observer's threshold. This measure of absolute efficiency was significantly correlated with the measure based on the cross-correlation between CIs and the ideal template in each age group (younger: $r = 0.75$, $p = .006$; older: $r = 0.72$, $p = .009$, both one-tailed) and overall ($r = 0.89$, $p < .001$, one-tailed). Consistent with our previous analyses, older adults had lower observed absolute efficiency than younger adults ($t(10.3) = 4.02$, $p = .001$, one-tailed; $M_Y = 1.77\%$, $M_O = 0.39\%$). These results suggest that the classification images were sensitive to the group and individual differences that affected absolute efficiency in our task.

Finally, Table 2 contains a measure of cross-session consistency, which was calculated by cross-correlating the z-scores of the smoothed images obtained in sessions 1 and 2. Consistency was significantly lower in older adults compared to younger adults, ($t(16.74) = 2.04$, $p = .029$, one-tailed; see also Fig. 2). In fact, older adults showed essentially *no* correlation at the group level ($t(9) = 0.11$, $p = .459$, one-tailed), and many showed a *negative* correlation between the sessions.

The data presented in Table 2 and Fig. 6 suggest that between-subject variation was larger in the older group. For example, averaged across sessions, GJW had the highest cross-correlation and lowest threshold, yet the classification image for that observer had no apparent structure and few significant pixels. In addition, two older adults with relatively high thresholds (BAG & GEZ) had classification images that were qualitatively similar to those of younger adults, and showed the greatest number of significant pixels, but still had relatively low cross-correlations.

4. Discussion

The current study used the sub-sampling variant (Nagai et al., 2013) of the classification image technique to compare the spatial sampling



(a) younger (b) older (c) overlap

Fig. 5. Spatial location and density of significant pixels, after 2900 trials, for younger (a) and older (b) observers, showing the number of observers relying on a given pixel location. Panel (c) illustrates the overlap between groups by highlighting only pixels that were statistically significant in the classification images from at least one younger and one older subject, colour coded by older participant. Instances where subject initials are not listed indicate either an absence of significant pixels, or significant pixels all fell within a region of overlap.

Table 2
Normalized cross-correlation with ideal template (absolute efficiency), root mean squared (RMS) signal contrast threshold (threshold), and number of significant ($p < .001$) pixels in the smoothed classification image for younger and older observers, for each session alone (S1, S2) and combined (both). Cross-session consistency represents the cross-correlation of the spatial distribution of z-scores in the smoothed classification images between S1 and S2.

Observer	Cross-Correlation			Threshold			# pixels			Cross-session consistency
	S1	S2	Both	S1	S2	Both	S1	S2	Both	
Younger										
BTP	0.05	0.10	0.11	0.04	0.04	0.04	26	29	73	0.42
EAW	0.09	0.09	0.12	0.05	0.05	0.05	29	4	35	0.34
JNW	0.05	0.02	0.05	0.09	0.07	0.08	9	34	1	−0.40
LKC	0.04	0.07	0.08	0.08	0.09	0.08	62	21	97	0.32
MVP	0.13	0.09	0.16	0.03	0.03	0.03	6	32	53	0.15
MXS	0.06	0.05	0.08	0.06	0.06	0.06	8	42	79	0.24
NOA	0.03	0.08	0.08	0.05	0.05	0.05	62	51	23	−0.03
RMG	0.07	0.10	0.12	0.06	0.07	0.07	19	12	69	0.42
SEC	0.10	0.11	0.14	0.04	0.04	0.04	58	12	91	0.39
TJP	0.06	0.05	0.08	0.05	0.06	0.06	42	53	84	0.30
mean:	0.07	0.08	0.10	0.06	0.06	0.06	32.1	29.0	60.5	0.21
Older										
BAG	0.04	0.00	0.03	0.11	0.09	0.10	50	22	51	0.34
BOC	0.03	0.03	0.04	0.09	0.07	0.08	9	2	1	0.16
GEZ	0.00	0.06	0.04	0.17	0.14	0.16	15	15	74	0.20
GJW	0.05	0.07	0.09	0.07	0.07	0.07	5	29	7	−0.18
HUG	0.03	0.01	0.03	0.11	0.10	0.10	35	10	15	−0.03
LON	−0.02	0.05	0.02	0.27	0.19	0.23	3	13	2	−0.13
LOR	0.00	0.04	0.02	0.26	0.26	0.26	14	0	3	0.13
MMM	−0.01	0.04	0.02	0.17	0.14	0.16	91	1	12	−0.20
MOF	0.02	0.06	0.05	0.08	0.11	0.09	4	30	7	−0.21
MRF	0.05	0.01	0.04	0.10	0.13	0.11	24	43	27	−0.01
mean:	0.02	0.04	0.04	0.14	0.13	0.14	25.0	16.5	19.9	0.01

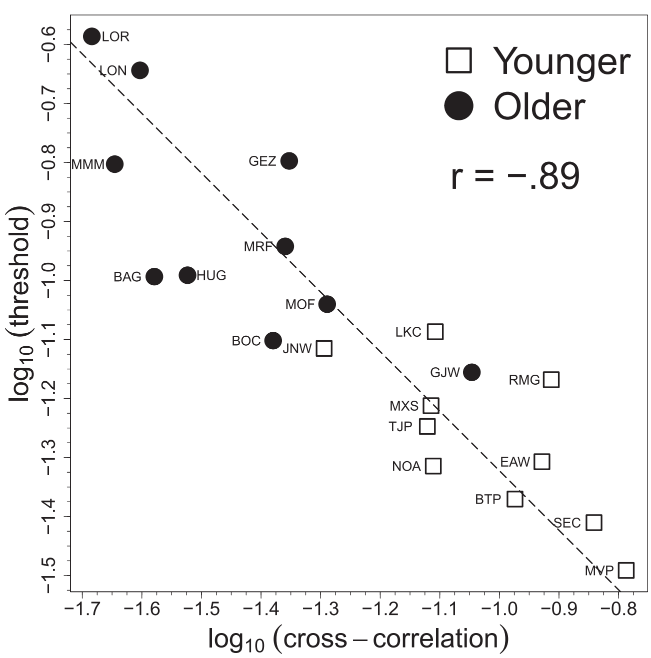


Fig. 6. Relationship between log-transformed root mean squared (RMS) contrast threshold and log-transformed normalized cross-correlation of the compressed raw classification image with the ideal template. Data combined across sessions. White squares and black circles represent younger and older observers, respectively. Dashed line represents regression line fit to data of both younger and older observers.

strategies used by younger and older observers performing a face discrimination task. Consistent with previous findings, classification images (CIs) from younger adults exhibited significant spatial structure, which suggests that they consistently relied most heavily on the eye/brow region when making their decisions. By contrast, most older observers showed no clear structure in their classification images, and demonstrated greater within-group and cross-session variability. In addition, older adults were less sensitive (higher contrast thresholds), less efficient (lower cross-correlation with ideal template), and had higher cross-session variability (lower correlation between z-scored CIs) compared to younger adults.

What does the lack of consistent structure in older adults' classification images suggest? One possibility is that our method lacked a sufficient number of trials to detect older observers' strategy. Relatedly, it could be that our method does not adequately capture the strategy used by older observers. The structure in a classification image reflects all aspects of a linear observer's visual processing, but may fail to capture non-linear processes (Murray et al., 2005). If non-linearities play a more significant role in perceptual processing for older adults than for younger adults, one would expect to see less obvious structure in CIs from older observers. One way which non-linearities might manifest themselves is that older adults might be more reliant on holistic processing (Adduri & Marotta, 2009; Creighton et al., 2012; Daniel & Bentin, 2012; Konar et al., 2013), which might not be adequately represented by the structure within classification images. However, one aspect of our findings is inconsistent with this hypothesis. Specifically, we found that a measure of absolute efficiency derived from classification images was significantly correlated with discrimination thresholds in both age groups, suggesting that classification images were capturing important aspects of the perceptual strategies used by both younger and older adults.

The model we use here assumes performance is constrained by two factors intrinsic to the observer: (1) internal noise and (2) calculation efficiency (Pelli & Farell, 1999). The value of the cross-correlation of an observer's classification image with the ideal template, which is associated with absolute efficiency (Murray et al., 2005), is affected by internal noise *and* calculation efficiency. In other words, cross-correlations will be highest – and absolute efficiency will be highest – for observers with low internal noise and who use templates that are similar to the ideal template (i.e., have a high *calculation* efficiency). Therefore, higher levels of internal noise in older adults might account for the lack of observable structure in older adults' classification images, the low cross-correlations with the ideal template, and the increased inter-session variability among older observers. There is some neurophysiological evidence that visual cortical neurons exhibit greater noise in older animals (Schmolesky, Wang, Pu, & Leventhal, 2000); however, other factors may contribute to an age difference in internal noise. For example, our results also are consistent with the possibility that older observers have greater variability in their response strategy across trials compared to younger adults, and this lack of consistency could result in reduced CI structure as well as higher estimates of internal noise (Burgess & Colborne, 1988) in the older group. Higher estimates of internal noise in older, relative to younger, observers have been found to contribute to age differences in motion detection (Bennett, Sekuler, & Sekuler, 2007), direction identification (Bennett et al., 2007; Bogfjellmo, Bex, & Falkenberg, 2013), orientation discrimination (Allard, Renaud, Molinatti, & Faubert, 2013; Betts, Sekuler, & Bennett, 2007), and detection of sine wave gratings at high (6–10 cpd) spatial frequencies (Pardhan, 2004). Thus, it is plausible that age differences in internal noise contributed to age differences in face discrimination thresholds found in the current study. It would therefore be fruitful for future studies to estimate the relative contributions of calculation efficiency and internal noise to older adults' reduced ability to discriminate faces. Another important consideration is how using the CI approach to age-differences in face perception might generalize to larger stimulus sets and/or different categories of faces (e.g., gender, age, race, etc.) than the two Caucasian male faces used here.

Nagai et al. (2013) found that classification images measured in a face discrimination task suggest that face processing strategies among ASD individuals fall into two distinct categories. Specifically, some ASD observers yielded classification images that were similar to those obtained from neurotypical observers, which suggests that they relied on information conveyed by pixels near the eyes and eyebrows. However, other ASD observers had classification images with atypical structure, suggesting that they based their responses on information conveyed by pixels on the forehead. Our results are similar, in the sense that they show that classification images in some, but not all, older adults are similar to those obtained in younger adults: two older observers showed a reliance on the eye/brow region similar to younger observers, whereas others relied on less informative areas such as the forehead or mouth. Our results also support the idea of greater within-group variability in older adults as we observed: (1) greater variability across sessions, and (2) greater absolute efficiency and lower contrast thresholds did not necessarily correspond with clearer structure, and spatial sampling strategies similar to younger observers did not necessarily correspond with higher absolute efficiency and lower contrast thresholds compared to older adults lacking structure.

An advantage of the classification image technique is that it makes no *a priori* assumptions about the spatial strategy employed by observers, and has a higher spatial resolution compared to other measures traditionally used in studying age-related changes in face perception. For example, studies of face perception that use eye-tracking often predefine regions of interest, such as the upper/lower halves of faces, or distinct facial features (most typically the eyes, nose, and mouth). By contrast, the response classification method examines performance at the level of a single pixel (or 2 × 2 clusters of pixels in the sub-sampling method), allowing us to observe older adults' reliance on regions such as the forehead, or cheek – areas typically assigned to an "other" category in eye-tracking studies. Furthermore, while it often is assumed that fixated positions represent regions critical to the observer performing the task, (i.e., information upon which they base their decisions and drives their performance), this may not always be the case. For example, Firestone et al. (2007) measured the gaze patterns of younger and older adults viewing faces who were subsequently given a surprise old/new memory recognition task. Despite increased sampling of face regions (more fixations and transitions between inner features such as the eyes), older adults demonstrated poorer recognition performance. The authors explain older adults' increased sampling as an attempt to use a compensatory strategy to overcome deficits in feature-binding (i.e., configural processing), yet their worse memory recognition suggests "such information was not used to make the recognition decision." Another study by Chan et al. (2011) further illustrates this point. In this study, younger and older adults were yoked to a gaze-contingent window tracking the patterns of either younger or older observers. Yoking older adults' eye-movements to those of younger observers did not increase performance on an old/new recognition task, and yoking younger adults' eye-movements to those of older observers did not decrease their performance. These studies thus lend support to the idea that the regions of the face that are sampled by older adults' eye movements may not correspond with the information actually driving older observers' decisions. The response classification technique, on the other hand, provides a direct measure of the relationship between information used and response accuracy.

In summary, the present study used a variant of the response classification technique to demonstrate both qualitative and quantitative differences in older and younger adults' performance on a face discrimination task. We replicated the finding that younger adults rely heavily on pixels in the eye/brow region when discriminating faces; however, the classification images from most older adults lacked obvious spatial structure. Nevertheless, we found that face discrimination thresholds in both age groups were strongly correlated with the similarity between the observed classification images and the ideal template, a result that suggests that classification images were sensitive to important aspects of face processing in both younger *and* older adults. Additionally, our results suggest that observer consistency may be lower, and between-subject variability may be higher, in older than younger adults.

Acknowledgements

This research was supported by grants from the Canadian Institutes for Health Research to PJB & ABS, and the Canada Research Chair Program to PJB. The authors kindly thank Donna Waxman for her assistance in data collection.

References

Adduri, C. A., & Marotta, J. J. (2009). Mental rotation of faces in healthy aging and alzheimer's disease. *PLoS One, 4*(7) e6120.
Allard, R., Renaud, J., Molinatti, S., & Faubert, J. (2013). Contrast sensitivity, healthy aging and noise. *Vision Research, 92*, 47–52.
Bennett, P. J., Sekuler, R., & Sekuler, A. B. (2007). The effects of aging on motion detection and direction identification. *Vision Research, 47*(6), 799–809.
Betts, L. R., Sekuler, A. B., & Bennett, P. J. (2007). The effects of aging on orientation discrimination. *Vision Research, 47*(13), 1769–1780.
Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ, 310*(6973), 170.
Bogfjellmo, L.-G., Bex, P. J., & Falkenberg, H. K. (2013). Reduction in direction discrimination with age and slow speed is due to both increased internal noise and reduced sampling efficiency. *Investigative Ophthalmology & Visual Science, 54*(8), 5204–5210.
Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition, 34*(4), 854–864.
Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.
Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A, 5*(4), 617–627.
Chaby, L., Narme, P., & George, N. (2011). Older adults' configural processing of faces: Role of second-order information. *Psychology and Aging, 26*(1), 71.
Chan, J. P., Kamino, D., Binns, M. A., & Ryan, J. D. (2011). Can changes in eye movement scanning alter the age-related deficit in recognition memory? *Frontiers in Psychology, 2*, 92.
Creighton, S. E., Sekuler, A. B., & Bennett, P. J. (2012). The effect of orientation and stimulus duration on older and younger adults' ability to identify facial expressions. *Journal of Vision, 12*(9) 964–964.
Daniel, S., & Bentin, S. (2012). Age-related changes in processing faces from detection to

identification: ERP evidence. *Neurobiology of Aging, 33*(1) 206–e1.

Dror, I. E., Schmitz-Williams, I. C., & Smith, W. (2005). Older adults use mental representations that reduce cognitive load: Mental rotation utilizes holistic representations and processing. *Experimental Aging Research, 31*(4), 409–420.

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., & Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision, 17*(14) 7–7.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap New York.* NY: Chapman and Hall.

Éthier-Majcher, C., Joubert, S., & Gosselin, F. (2013). Reverse correlating trustworthy faces in young and older adults. *Frontiers in Psychology, 4*, 592.

Firestone, A., Turk-Browne, N. B., & Ryan, J. D. (2007). Age-related deficits in face recognition are related to underlying changes in scanning behavior. *Aging, Neuropsychology, and Cognition, 14*(6), 594–607.

Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198.

Gaspar, C. M., Bennett, P. J., & Sekuler, A. B. (2008). The effects of face inversion and contrast-reversal on efficiency and internal noise. *Vision Research, 48*(8), 1084–1095.

Goffaux, V., & Dakin, S. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology, 1*, 143.

Goffaux, V., Poncin, A., & Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 year old). *PLoS One, 10*(9) e0138812.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research, 39*(21), 3537–3560.

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science, 28*(2), 167–207.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261–2271.

Habak, C., Wilkinson, F., & Wilson, H. R. (2008). Aging disrupts the neural transformations that link facial identity across views. *Vision Research, 48*(1), 9–15.

Halberstadt, J., Ruffman, T., Murray, J., Taumoepeau, M., & Ryan, M. (2011). Emotion perception explains age-related differences in the perception of social gaffes. *Psychology and Aging, 26*(1), 133.

Heisz, J. J., & Ryan, J. D. (2011). The effects of prior exposure on face processing in younger and older adults. *Frontiers in Aging Neuroscience, 3* 15–15.

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*(1), 38–43.

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision Research, 88*, 38–46.

Lee, Y., Grady, C. L., Habak, C., Wilson, H. R., & Moscovitch, M. (2011). Face processing changes in normal aging revealed by fmri adaptation. *Journal of Cognitive Neuroscience, 23*(11), 3433–3447.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.

Murphy, N. A., & Isaacowitz, D. M. (2010). Age effects and gaze patterns in recognising emotional expressions: An in-depth look at gaze measures and covariates. *Cognition and Emotion, 24*(3), 436–452.

Murray, J. E., Halberstadt, J., & Ruffman, T. (2010). The face of aging: Sensitivity to facial feature relations changes with age. *Psychology and Aging, 25*(4), 846.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision, 5*(2) 5–5.

Nagai, M., Bennett, P. J., Rutherford, M., Gaspar, C. M., Kumada, T., & Sekuler, A. B. (2013). Comparing face processing strategies between typically-developed observers and observers with autism using sub-sampled-pixels presentation in response classification technique. *Vision Research, 79*, 27–35.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699.

Obermeyer, S., Kolling, T., Schaich, A., & Knopf, M. (2012). Differences between old and young adults' ability to recognize human faces underlie processing of horizontal information. *Frontiers in Aging Neuroscience, 4*, 3.

Pachai, M. V., Sekuler, A. B., & Bennett, P. J. (2013a). Masking of individual facial features reveals the use of horizontal structure in the eyes. *Journal of Vision, 13*(9), 411.

Pachai, M. V., Sekuler, A. B., & Bennett, P. J. (2013b). Sensitivity to information conveyed by horizontal contours is correlated with face identification accuracy. *Frontiers in Psychology, 4*(74), 1–8.

Pardhan, S. (2004). Contrast sensitivity loss with aging: Sampling efficiency and equivalent noise at different spatial frequencies. *JOSA A, 21*(2), 169–175.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A, 16*, 647–653.

Pilz, K. S., Bennett, P. J., & Sekuler, A. B. (2010). Effects of aging on biological motion discrimination. *Vision Research, 50*(2), 211–219.

R Development Core Team. (2016). R: A Language and Environment for Statistical Computing. URL http://www.r-project.org.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews, 32*(4), 863–881.

Schmolesky, M. T., Wang, Y., Pu, M., & Leventhal, A. G. (2000). Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. *Nature Neuroscience, 3*, 384–390.

Schwarzer, G., Kretzer, M., Wimmer, D., & Jovanovic, B. (2010). Holistic face processing among school children, younger and older adults. *European Journal of Developmental Psychology, 7*(4), 511–528.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402–409.

Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition, 27*(3), 538–552.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*(5), 391–396.

Sekuler, A. B., Pachai, M. V., Creighton, S. E., & Bennett, P. J. (2014). Age-related effects on selective processing of horizontal structure in whole-face context. *Journal of Vision, 14*(10) 573–573.

Slessor, G., Riby, D. M., & Finnerty, A. N. (2012). Age-related differences in processing face configuration: The importance of the eye region. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*(2), 228–231.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*(3), 184–189.

Sullivan, S., Ruffman, T., & Hutton, S. B. (2007). Age differences in emotion recognition skills and the visual scanning of emotion faces. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 62*(1), P53–P60.

Wong, B., Cronin-Golomb, A., & Neargarder, S. (2005). Patterns of visual scanning as predictors of emotion identification in normal aging. *Neuropsychology, 19*(6), 739.

Yu, D., & Chung, S. T. (2011). Critical orientation for face identification in central vision loss. *Optometry and Vision Science: Official Publication of the American Academy of Optometry, 88*(6), 724.