

SCIENTIFIC REPORTS



OPEN

Integration of facial features under memory load

K. Ölander, I. Muukkonen, T. P. Saarela & V. R. Salmela 

Received: 13 July 2018
 Accepted: 11 December 2018
 Published online: 29 January 2019

Simple visual items and complex real-world objects are stored into visual working memory as a collection of independent features, not as whole or integrated objects. Storing faces into memory might differ, however, since previous studies have reported perceptual and memory advantage for whole faces compared to other objects. We investigated whether facial features can be integrated in a statistically optimal fashion and whether memory maintenance disrupts this integration. The observers adjusted a probe – either a whole face or isolated features (eyes or mouth region) – to match the identity of a target while viewing both stimuli simultaneously or after a 1.5 second retention period. Precision was better for the whole face compared to the isolated features. Perceptual precision was higher than memory precision, as expected, and memory precision further declined as the number of memorized items was increased from one to four. Interestingly, the whole-face precision was better predicted by models assuming injection of memory noise followed by integration of features than by models assuming integration of features followed by the memory noise. The results suggest equally weighted or optimal integration of facial features and indicate that feature information is preserved in visual working memory while remembering faces.

Faces contain socially important information and consequently our visual system is very sensitive locating and detecting faces¹ and a large network of brain areas is specialized in processing of faces^{2,3}. Facial information can be roughly divided into two categories: changeable features, such as emotional expressions, and invariant features, such as identity. Previous studies suggest that changeable and invariant features are, at least partly, processed by different mechanisms⁴. Monkey single cell studies⁵ and human fMRI studies⁶ have shown evidence for norm-based coding of identities, suggesting that identities are represented in a multidimensional facial feature space⁷. Different facial features, however, contribute differently for face perception; the regions around mouth and eyes are most informative^{8,9} and the discriminability of head shape and hair-line is better than discriminability of mouth, eyes and eyebrows¹⁰.

A common notion in the literature of face perception is that of holistic or configural processing^{11–14}, which suggests that the perception of a whole, upright face is different from its parts or faces presented upside-down. The whole face benefit might be due to an optimal integration of facial features. Previously, statistically optimal integration has been studied in face recognition by measuring contrast thresholds^{15,16}. In these studies contrast thresholds for identity recognition were measured using facial features in small circular apertures, and compared to the contrast threshold of the whole face (i.e., all features presented at the same time). The results suggested that facial feature integration is optimal^{15,16} or supra-optimal when spatial uncertainty to feature locations was added¹⁷. Optimal integration of facial form and motion cues was also found in identity matching task with high contrast synthetic faces¹⁸.

The representations of faces, and visual objects in general, can deteriorate due to several factors. Recognition may be impaired due to visual clutter or noise, holding visual representations in memory for prolonged time, or trying to remember several objects at the same time. According to current understanding of visual working memory, there is a trade-off between the memory capacity and precision^{19–23}, that is, the more objects we try to remember, the less precise the memory representations are. For primary visual features, the decline in memory precision due to multiple items in memory can be explained by an increasing noise during memory maintenance. For complex items containing multiple features, the memory noise can have several effects. If the individual features are bound together to form object representations²⁴, the memory noise could also affect the binding or integration, in addition to the features. Previous studies suggest independent storing of different features of simple^{25–27} and complex visual objects²⁸. Upright faces, however, are remembered better than other complex visual objects²⁹

Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland. T. P. Saarela and V. R. Salmela contributed equally. Correspondence and requests for materials should be addressed to V.R.S. (email: viljami.salmela@helsinki.fi)

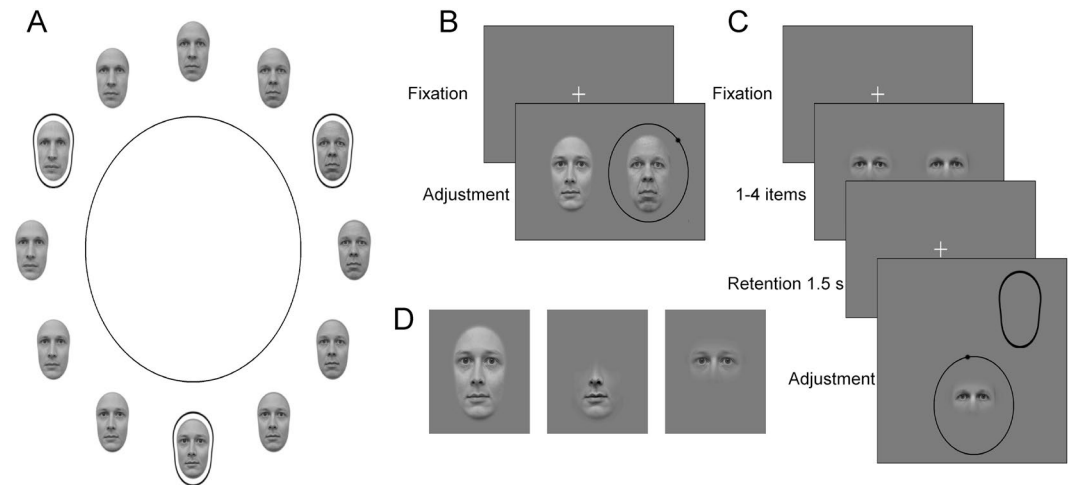


Figure 1. Stimuli and experimental setup. (A) Circular identity spaces were created by morphing three original identities with neutral expression to each other (originals marked with a black outline). (B) In perceptual matching experiment the observers matched to stimuli presented side-by-side (a black dot moved around the black circle for visual feedback). (C) In the memory experiment, 1–4 stimuli were presented and observers adjusted the identity of a probe item to match the identity of a cued item (oval outline). (D) All the experiments were conducted with whole faces, isolated mouth (lower part of the face) and isolated eyes (upper part of the face). The identities in the figure are examples, not the ones that were used in the actual experiment.

or faces presented upside-down³⁰. Thus, for images of human faces, memory noise could corrupt the whole, integrated face representation, it could have different effects on different features if they are independently stored, or it could disrupt the integration or binding as such.

We studied integration of facial features using high-contrast images of real faces and a task where the observer adjusted the identity of a probe face or an isolated feature to match a target face/feature (Fig. 1A,D). The stimuli were presented side-by-side (perception; Fig. 1B) or sequentially with a retention period in between the target and probe (memory; Fig. 1C). In the memory condition we varied memory load and the observers had to memorize one to four identities. Precision of adjustment was estimated by fitting a wrapped Cauchy distribution to the distribution of adjustment errors. We computed predictions for the whole-face stimulus in the perception condition assuming a model observer who optimally integrates the two features (eye and mouth region). Human observers' performance was well predicted by this model. We then predicted the performance for the whole face stimulus in the memory conditions assuming that the whole face is affected by memory noise in the same way the individual features are. We tested several models, assuming that either faces are stored as integrated objects or that features are stored separately, and assuming that either most reliable cue is used, cues are equally weighted or optimally integrated. The human data was best predicted by models assuming optimal or equally weighted integration of separately stored features.

Results

Precision of adjustments. In the perception experiment, the observers viewed two stimuli presented side-by-side and adjusted the stimuli on the right side to match in identity the stimuli on the left side (Fig. 1B). Although the stimuli were constantly visible, the perceptual matching of identity was not perfect. Instead, the adjustment errors formed distributions, which were well fit by a wrapped Cauchy distribution for each individual observer (data and fits for observer 6 shown in Fig. 2, perception data in the first column). The error distributions were narrower for the whole face than for the eyes and mouth stimuli (Fig. 2, first, second and third rows, respectively). The same pattern of results was found in the average data (Fig. 3, first column).

The mean \pm *s.e.m* circular standard deviation of perceptual errors was 0.44 ± 0.049 , 0.56 ± 0.084 , and 0.56 ± 0.082 radians for faces, eyes, and mouth, respectively. The width of the error distribution was narrower for whole faces than for eyes (one-sided $t(7) = 2.67$, $p = 0.016$; $BF_{10} = 5.17$) and mouth (one-sided $t(7) = 3.13$, $p = 0.008$; $BF_{10} = 8.64$), while no difference between the eyes and mouth was found ($t(7) = 0.035$, $p = 0.973$; $BF_{10} = 0.336$). The corresponding mean \pm *s.e.m* concentration ρ values of the fitted wrapped Cauchy distribution, were 0.81 ± 0.022 , 0.77 ± 0.037 and 0.77 ± 0.041 , for faces, eyes and mouth, respectively. The concentration parameter was higher for whole faces than for eyes (one-sided $t(7) = 2.20$, $p = 0.032$; $BF_{10} = 3.020$) and mouth (one-sided $t(7) = 1.96$, $p = 0.046$; $BF_{10} = 2.288$), but no difference was found for isolated eyes and mouth ($t(7) = 0.070$, $p = 0.946$; $BF_{10} = 0.337$). The Bayesian tests provided confirmatory evidence in favor for our hypotheses (all $BFs > 2.8$), and in favor for *null hypotheses* when isolated features were compared ($BFs < 0.34$). Thus the perceptual precision was better for the whole face than isolated features, but the precision for features did not differ from each other.

To further test the difference between the features and faces, two additional fits of Cauchy distributions were made. The first model – reported above – contained separate parameters for all stimulus types. The second model contained different parameters for features and faces, and the third model contained same parameter for all

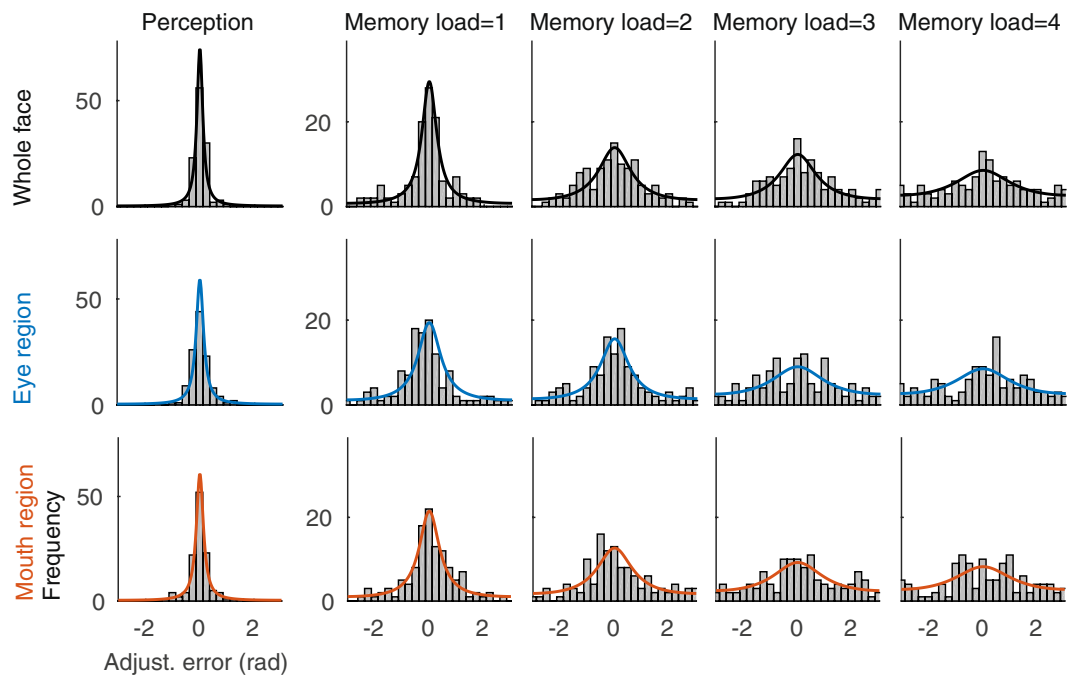


Figure 2. Error distributions of an individual observer (observer 6). The adjustment error distributions (gray bars) and fitted wrapped Cauchy distributions, appropriately scaled (black, blue, and red lines for whole face, eye region and mouth region, respectively).

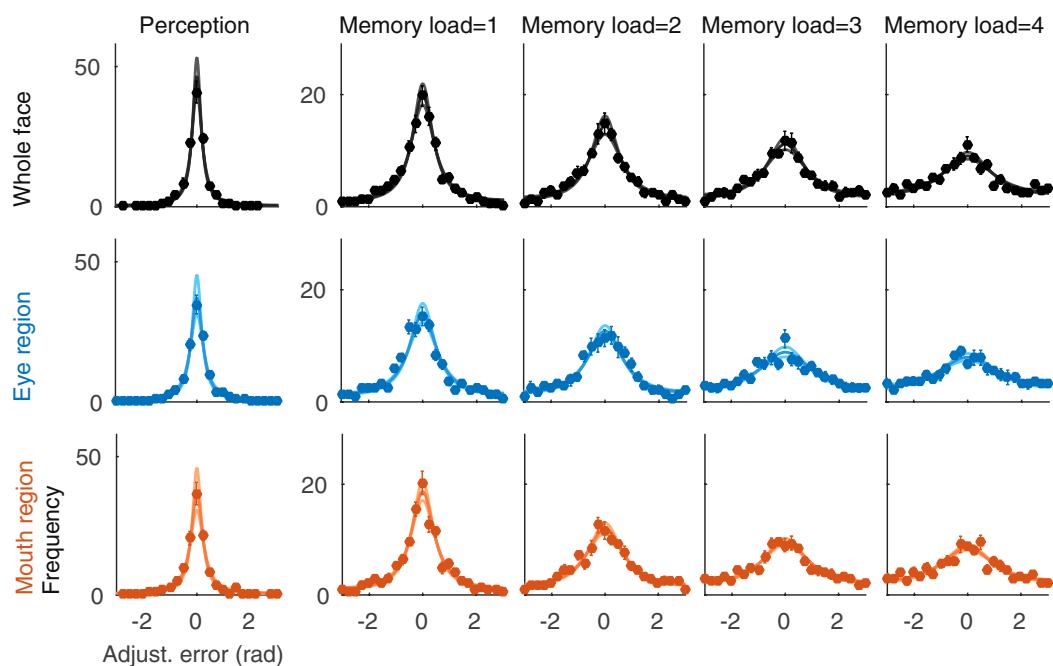


Figure 3. Error distributions averaged over all observers. The average error distributions (circles) and average fitted wrapped Cauchy distributions, appropriately scaled (black, blue, and red lines for whole face, eye region and mouth region, respectively). Error bars depict standard error of mean.

stimulus types. Better fits were obtained with models one and two than model three. There was some individual variability, but in comparison to model three, the AIC scores were smaller for models one and two, mean \pm *s.e.m* were 0.39 ± 1.32 (sum over observers 3.12) and 1.21 ± 1.50 (sum over observers 9.65), respectively. Since the precision for eyes and mouths was so similar, separate parameters for the different features (model one) did not provide additional benefit (AIC difference to model two was -0.82 ± 0.53 , sum over observers -6.52).

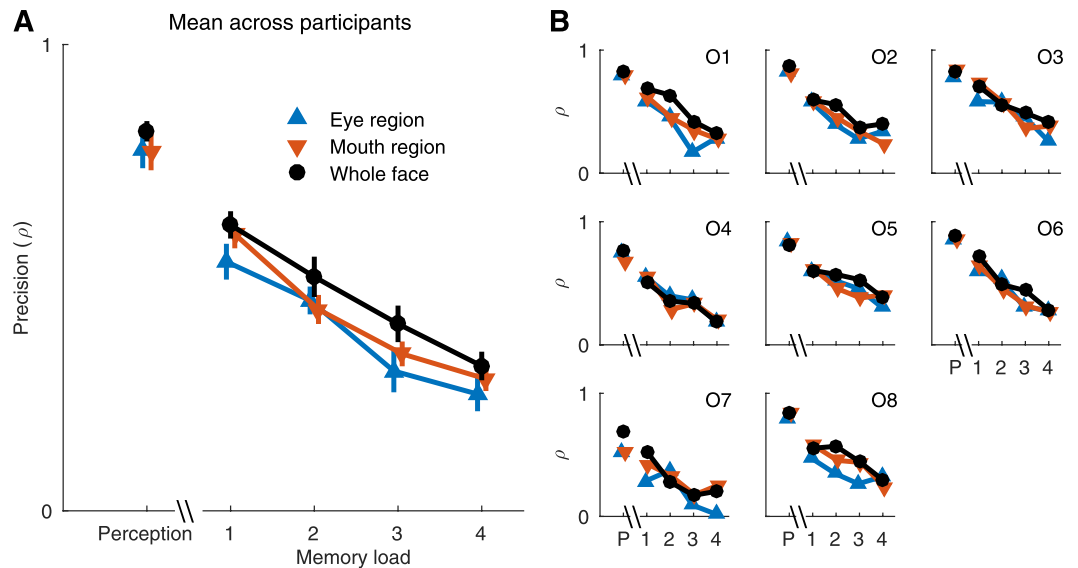


Figure 4. Results from both experiments. Measured precision as the concentration parameter of fitted wrapped Cauchy distributions. (A) The mean \pm s.e.m, across all observers. (B) The precision of all observers. Color conventions as in Fig. 2. P = Perception.

In the second experiment a 1.5 second memory period was added between the target and probe stimuli, and the observers' task was to adjust the probe to match the target identity in memory (Fig. 1C). Consequently, the width of the adjustment error distribution increased. This was evident already with just one stimulus to be remembered (Fig. 2/3, second column). Adding the retention interval to the task doubled the averaged standard deviation of errors to 0.92 ± 0.059 , 1.09 ± 0.073 , and 0.97 ± 0.048 radians for faces, eyes and mouth, respectively. When the memory load was increased from one to four (Fig. 2/3, columns 2–5), the width of the error distributions further increased and the mean \pm s.e.m standard deviation of errors for four items was 1.56 ± 0.078 , 1.62 ± 0.055 , 1.60 ± 0.062 radians for faces, eyes and mouth, respectively). However, even with the largest memory load, the observers were able to memorize the stimuli and the error distributions were not flat (Fig. 2/3, column five).

The effect of memory load on the standard deviation of errors was statistically highly significant ($F(3,21) = 79.95$, $p < 0.001$; $\text{Log}(\text{BF}_{10}) = 54.73$) as well as the main effect of stimulus type ($F(2,14) = 7.83$, $p = 0.005$; $\text{BF}_{10} = 44.77$). There was, however, no interaction between the stimulus type and memory load ($F(6,42) = 1.32$, $p = 0.272$; $\text{BF}_{10} = 0.269$) suggesting that the memory precision declined similarly for individual features and the whole face. Similar results were found for the concentration parameter of the fitted distributions, main effects of load ($F(3,21) = 97.95$, $p < 0.001$; $\text{Log}(\text{BF}_{10}) = 60.470$) and stimulus type ($F(2,14) = 10.637$, $p = 0.002$; $\text{Log}(\text{BF}_{10}) = 6.411$), but no interaction ($F(6,42) = 0.887$, $p = 0.513$; $\text{BF}_{10} = 0.171$). The low BF in the last (0.171) test provides quite strong evidence against the interaction of load and stimulus type.

When fitting additional models to the memory data, better fits were obtained with models that contained a separate concentration parameter for features and whole face, except for highest memory load (mean \pm s.e.m AIC differences between third and second model were 0.27 ± 0.94 , 1.48 ± 1.48 , 0.08 ± 0.63 , and -1.27 ± 0.38 (sum over observers: 2.16, 11.86, 0.68, and -10.13); AIC differences between third and first model were 0.03 ± 1.38 , 0.41 ± 1.53 , -0.21 ± 1.10 , and -1.61 ± 0.96 (sum over observers: 0.24, 3.3, -1.64 , and -12.85)). And again due to similar concentration for eyes and mouth, separate parameters for eyes and mouth did not provide benefit (AIC difference between second and first model were -0.24 ± 0.96 , -1.07 ± 0.31 , -0.29 ± 0.69 , and -0.34 ± 0.81 (sum over observers: -1.92 , -8.55 , -2.32 , and -2.72)).

One possible explanation for the difference between the features and the whole faces is that the observers enjoyed the whole face condition more and therefore spent more time in adjusting the probe in the whole face condition. However, the Bayesian test revealed evidence *against* an effect of condition (eyes vs. mouth vs. face) on adjustment durations in perception ($F(2,14) = 0.752$, $p = 0.489$; $\text{BF}_{10} = 0.381$) or in memory experiment ($F(2,14) = 2.126$, $p = 0.156$; $\text{BF}_{10} = 0.313$). In the memory experiment, memory load had no main effect on durations ($F(3,21) = 0.559$, $p = 0.648$; $\text{BF}_{10} = 0.095$), and no interaction with stimulus type ($F(6,42) = 0.315$, $p = 0.925$; $\text{BF}_{10} = 0.094$). In the latter tests, the evidence against any effects is particularly strong (BF less than 0.1), indicating that observers in all condition made similar effort in adjusting the probe stimulus. In perception experiment, the mean \pm s.e.m durations for faces, eyes, and mouths were 11.0 ± 0.81 , 10.3 ± 1.32 , and 11.7 ± 1.51 seconds, respectively. In memory experiments, observers spent much less time adjusting the probe, the mean \pm s.e.m durations for faces, eyes, and mouths were 6.3 ± 0.68 , 6.6 ± 0.72 , and 6.2 ± 0.73 seconds.

To summarize all the results, the average and individual precision (i.e., concentration parameter ρ of the fitted wrapped Cauchy distributions) in perception and memory experiments are shown in Fig. 4A,B, respectively. Perceptual precision was substantially better than memory precision, as expected. On average, precision was better for the whole face than for the isolated features in every condition (Fig. 4A), although there was some

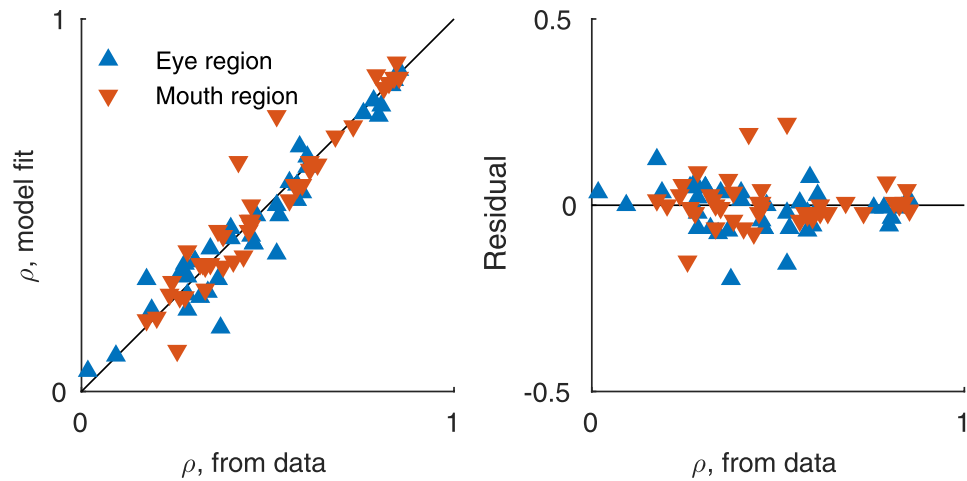


Figure 5. Fit and measured precision values, all observers and memory loads. See text for details.

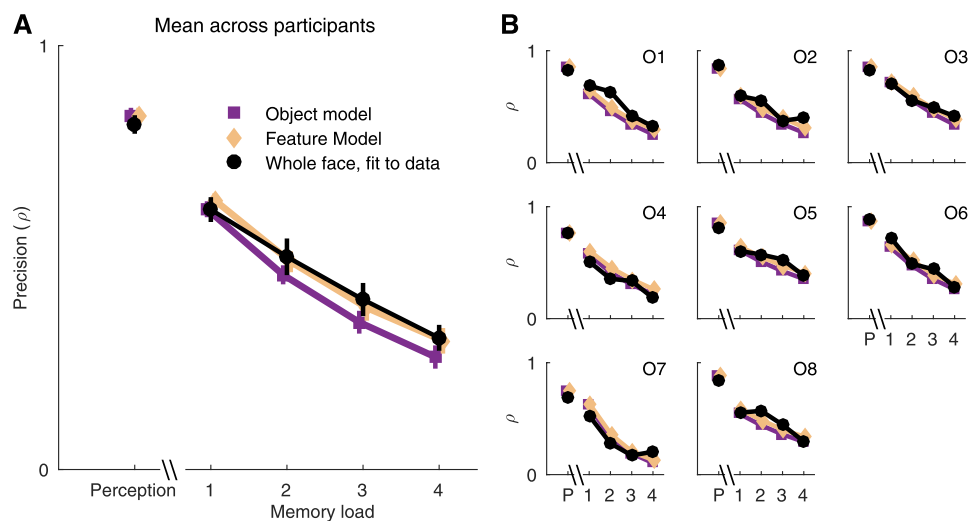


Figure 6. Predicted and measured precision values for the whole face. (A) The average precision across all observers. Feature Model predicted the measured whole face precision most accurately. (B) Precision values of each observer. The accuracy of both models varied across observers, overall Feature Model predicted the precision better than the Object Model. P = Perception.

individual variation (Fig. 4B). For 7/8 observers the perceptual precision and for 5–7/8 observers the mnemonic precision was better for the whole face than for isolated features, depending on the memory load (Fig. 4B).

Modeling. Since the precision of the isolated features did not reach the precision of the whole face (Fig. 4A), it seems that the facial features are integrated when matching facial identities of the whole face. To quantify the feature integration, we predicted the whole face performance based on the performance with isolated features and assuming optimal integration. We assumed that three independent sources of noise (feature, memory delay, and load) limit precision. We estimated the precision of each type of noise from the isolated feature data (eyes and mouth), separately for each observer (see Methods). The fits captured the effects of memory on precision well, and the estimated precision in each condition and for every observer was very close to the observed data (Fig. 5).

We then tested for integration of features in the whole-face condition. We devised two versions of an optimal-integration model, in which memory noise corrupts perceptual precision before (*Feature Model*) or after (*Object Model*) the integration of features, and compared human observer performance against model predictions. In addition, we fitted two versions of an equal-weighting-of-cues model, in which the cues were integrated before or after memory noise, and a most-reliable-cue model. The predictions of the optimal-integration models are shown in Fig. 6, along with the whole-face data replotted from Fig. 4. The perception-only (no memory load) prediction is identical for the two models, and especially the average data (Fig. 6A) is in accordance with this prediction. Both models predict a similar and systematic decline in performance as a function of memory load, but the Feature Model, in which memory noise corrupts precision *before* integration, predicts better performance overall in the memory conditions. Averaged over observers, the Feature Model also more accurately

predicts performance. The individual observer data (Fig. 6B) are noisy, but based on likelihood ratios, the Feature Model better predicted performance for 6 out of 8 observers. When comparing all fitted models, for 6/8 observers, higher likelihoods were found for models containing integration of features after memory noise rather than before memory noise. For 3/8 observers the best model contained optimal integration, for 3/8 observers the best model was equal weighting, and for two observers the most reliable cue provided the best fit for data. The predictions of the optimal-integration and equal-weighting-of-cues models were very similar.

Discussion

We investigated integration of facial features in a face identity matching task, and tested whether adding a retention period and an increasing memory load disrupts feature integration in memory. To this end, we measured adjustment errors for facial features and the whole face while observers matched stimuli presented side-by-side as well as stimuli separated by a memory interval. The perceptual and mnemonic precision for the whole face was better than for the isolated features, and the precision decreased due to retention period and memory load. Memory precision for the whole face was well predicted by models with late integration of noisy features. Alternative models – early integration of features followed by an additive memory noise – underestimated memory precision. We compared optimal integration with equal weighting of cues and choosing most reliable cue. The results suggest that facial features are integrated in statistically optimal fashion or by equally weighting features, and that feature information is preserved while memorizing complex objects.

A growing amount of evidence suggests flexible resources in visual working memory^{31,32}. According to current understanding, we can trade precision with capacity, that is, remember only few items very precisely or larger amount of objects with less resolution. This can be achieved either by continuous resources^{21,33}, variable precision^{19,20} or by averaging discrete memory representations for improved precision²³. Most studies on the precision of working memory have used simple visual features or shapes as stimuli to be memorized. With more complex or naturalistic stimuli, such as images of human faces, the precision of memory representations can decrease in several ways. We can remember complex objects either as a whole object, or as a collection of features combined with binding information^{24,34}.

Memory binding is typically studied with combining simple visual features, such as color and orientation. Previous studies suggests that memory loads affects binding²⁵ and emphasize feature locations, that is, features are bound mainly due to the shared location^{35,36}. Features of simple visual objects are independently stored^{25–27}. Similarly, complex, real-world objects are not remembered as unitary or bound objects since the memorability of features are partly independent²⁸. Our results with faces suggest that memory noise similarly degrades both the isolated features as well as the integrated objects, that is, there was no interaction between the memory load and stimulus type. Furthermore, predicting the memory precision for the whole face with different models favored the models in which memory noise was added before feature integration. This suggests that unfamiliar faces are stored in memory as collection of features or that feature information is preserved while storing facial identity in working memory. Interestingly, storing a face as an integrated object is not the most optimal strategy since – according to our modeling – storing features separately predicted better memory performance. Although there was some variability, majority of our observers seemed to utilize this strategy.

It has been suggested that visually complex objects employ more working memory resources than visually simple objects³⁷. In our experiment the whole-face stimuli were remembered more precisely than the single-feature stimuli, although faces are visually more complex. However, the precision declined similarly for both whole face and isolated features. There are several advantages for faces compared to other visual objects. Faces are remembered better than other objects, if the encoding time is long enough²⁹, and more precisely than line orientations when multiple items need to be remembered³⁸. Face inversion effect has also been reported during memory maintenance, that is, memory precision for upright faces is better than for upside-down faces³⁰. Our results suggest that the memory advantage for faces is not due to storing faces as integrated objects.

The difference in precision between facial features and the whole faces was not as large as we expected on the basis of holistic or efficient processing of whole upright faces. Furthermore, the precision for mouth and eyes was quite similar in all conditions. These could be due to our feature stimuli; the whole face was split in half and we compared identity information in the lower and upper part of the face. The whole face benefit, as well as the difference between the features, would likely be larger if facial features would be chosen with a smaller aperture, i.e., just a left eye, or an eye without the eyebrow. Furthermore, the reliability of different cues could be varied with varying aperture size, which in turn would make predictions of optimal integration and equal weighting of cues more different. In the current data, we cannot separate these two models.

All the identities we used were unfamiliar to the observers. Identity processing differs between unfamiliar and familiar faces, and facial expertise is more pronounced with familiar faces³⁹. With familiar faces, observers would have had much more exposure to the stimuli and different memory representations, and they would likely have used different strategies or weighting of features while memorizing identities. However, as the main interest of the study was short-term or working memory, long-term memory representation of familiar faces could have affected these processes, as observers could have used other identity information, such as names, while memorizing faces.

Previously optimal integration of facial features has been shown for detection of faces at contrast threshold^{15–17} as well as discrimination identity of dynamic synthetic faces¹⁸. In accordance with these studies, our results suggest optimal integration or equal weighting of facial features in an adjustment task. In comparison to the previous integration studies, we added a retention period and varied the memory load. The memory maintenance or load did not have much effect on the integration as such. In conclusion, our results suggest optimal integration or equal weighting of facial features in identity matching task, and that faces are stored as collection of features, which are integrated during retrieval.

Methods

Observers. Eight observers (4 female, 22–31 years) with normal or corrected to normal vision and without known deficits in face perception participated in the experiment. All observers received a study credit for participating. All the experiments were conducted in accordance with the Declaration of Helsinki. A written informed consent was collected from the observers before the measurements, and the experiments were approved by the Ethics Review Board in the Humanities and Social and Behavioural Sciences of the University of Helsinki.

Stimuli. We chose 60 face images depicting different identities from the Radboud⁴⁰ and FACES⁴¹ databases. All faces had a neutral expression. Half of the identities were female and half male. Images were divided into groups of three and morphed from one identity to another by using Abrosoft FantaMorph software. As a result, 20 different circular identity spaces, which each contained 300 images in total, were formed. In each of these spaces the identity changed continuously between the three original identities (Fig. 1A). The gender did not change within the identity space.

The three different stimulus types, whole face, mouth region (lower part of the face) and eye region (upper part of the face), were obtained by applying Gaussian masks to the images (Fig. 1D). The masks were identical across all identities. The whole faces were first extracted from the original images with a mask, which shape was determined by two radial frequency components (first component: RF = 2, amplitude = 0.22, phase = 270°; second component: RF = 3, amplitude = 0.04, phase = 180°). The eye and mouth regions were extracted from the whole face with masks defined by a sum of Gaussians, i.e., the eye region was extracted with two Gaussians around the left and right eye and the mouth region with two Gaussians around the nose and mouth. The standard deviations of the Gaussian masks for the left and right eye were $\sigma_x = 1.6^\circ$ and $\sigma_y = 0.8^\circ$, for nose $\sigma_x = 0.6^\circ$ and $\sigma_y = 0.8^\circ$, and mouth $\sigma_x = 2.0^\circ$ and $\sigma_y = 1.2^\circ$. The size (width/height) of the whole face, eyes and mouth were $4.1^\circ \times 6.0^\circ$, $4.1^\circ \times 1.7^\circ$ and $3.7^\circ \times 3.2^\circ$, respectively. Gray-scale faces/features were displayed on a mid-gray background and the RMS-contrast of the stimuli was 0.19 ± 0.01 . All the image processing was conducted with Matlab.

Experiments were conducted in a dimly lit room. Stimuli were shown on a linearized VIEWPixx monitor (VPixx Technologies Inc., Canada). Observers sat 92 cm away from the monitor and their head rested on a chin-forehead stand. The viewing area extended 29.5×18.8 degrees. The stimulus presentation was controlled with Psychophysics Toolbox extension of Matlab⁴².

Procedure. The precision of perceiving and remembering facial identity was measured with the method of adjustment. Two experiments (perception and memory) were conducted using three different types of stimuli: (1) the whole face, (2) mouth region only, and (3) eye region only. In the perception experiment, a target stimulus and a probe stimulus were presented simultaneously, side-by-side (Fig. 1B). The observers' task was to adjust the identity of the probe to match the target. The target and the probe were always similar, i.e., whole face, or eye or mouth region. The target was always on the left side of the display and the probe on the right side of the display. The observer used up/down arrow keys for coarse adjustment (4.8° steps in the identity space) and left/right arrow keys for precise adjustment (1.2° steps in the identity space). When the observer was content with the adjustment, he/she initiated the next trial by pressing the spacebar. The maximal adjustment time was limited to 30 seconds. The positions of the target and the starting position of the probe stimulus in the identity space were random, except the probe was initially at least $\pm 30^\circ$ away from the target. For visual feedback on the identity space, there was a thin black circle around the probe stimulus, and a black dot moved along the circle according to observers' adjustment (Fig. 1B).

In the memory experiment, memory precision was measured while varying the memory load. First, 1–4 stimuli were shown and the observers' task was to memorize the identities of the stimuli. The stimuli were always shown on the same, fixed locations, and for 0.5 s per face, i.e., one face was shown for 0.5 s and three faces were shown for 1.5 s. After a 1.5 s retention period, a probe stimulus was presented on the bottom of the screen, and the observers' adjusted the probe stimulus to match the target identity, which was indicated with a spatial cue (outline of the face; Fig. 1C). The adjustments were done similarly as in the perception experiment. In the conditions with more than one stimulus, all of the stimuli were always from different identity circles.

All observers conducted the perception experiment first. For each experiment and condition, 120 trials were measured in two blocks of 60 trials. The order of the four memory loads and three stimulus types, and the two blocks in the memory experiment were randomized and balanced across observers. In every block, all of the 20 identity spaces were probed three times, once in each third of the circular space.

Data analysis. The adjustment error on each trial was obtained by computing the angle between the adjusted identity and the target identity in the circular identity space. To quantify the precision of the adjustments, we fit wrapped Cauchy distributions⁴³ to the adjustment error distributions. Wrapped Cauchy density function is given by:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)} \quad (1)$$

where θ is the angle, μ is a location parameter, and the concentration parameter ρ defines the precision, varying between 0 (uniform circular distribution) and 1 (distribution concentrated at μ). Fitting was done separately for each observer, condition, and stimulus type by numerically finding the maximum likelihood values for the parameters given the data. The mean on the distribution (location parameter μ) was set to zero. We used wrapped Cauchy instead of von Mises distribution because von Mises failed to capture the shape of the error distribution, especially when the error distribution was very concentrated. We quantified the difference in the goodness of fit between the wrapped Cauchy and von Mises by comparing their log-likelihoods. Average log-likelihood ratio for wrapped Cauchy vs. von Mises was 21.97 ± 7.97 , confirming that wrapped Cauchy gave a better fit. In addition,

we confirmed that the shape of the distributions did not differ between eyes, mouths and whole faces by fitting wrapped stable distribution⁴⁴ to the data with shape and concentration as free parameters. The shape parameter value 1 corresponds to a wrapped Cauchy distribution and value 2 corresponds to a wrapped Gaussian distribution. The mean \pm s.e.m shape values were 1.08 ± 0.048 , 1.14 ± 0.066 and 1.07 ± 0.091 for mouth, eyes, and face, respectively, and did not differ significantly from each other. All further analyses and modeling were done using wrapped Cauchy.

In contrast to some previous studies on visual memory, we used only a circular distribution, not a mixture of a circular and a uniform distribution, since the wrapped Cauchy alone gave a good fit for the data. Further, for the modeling (see below) it was essential to get a good estimate of precision in each experimental condition, and this becomes problematic with a mixture of a circular and a uniform distribution, since the spread of the observations could be absorbed either by the weight of the uniform distribution or the concentration parameter of the circular distribution, especially when the spread is large.

Statistical analyses were conducted with JASP software^{45,46}. Paired sample t-tests and repeated measures ANOVAs were used as well as Bayesian paired samples t-tests and Bayesian ANOVAs. The t-tests were two-sided unless otherwise noted. The Bayes Factors (BF) are reported relative to the null model and in two-way ANOVAs the BF are reported relative to the null model including the other effects.

Modeling. We developed a model to investigate whether facial features are optimally integrated for identification. The model assumes independent processing and statistically optimal integration of noisy features, and a further corruption of precision during memory maintenance by additional neural noise.

During the experiment, the observer adjusts one of the two stimuli so that the ‘perceptual distance’ between the two is minimized. The model observer does this using a noisy decision variable r , which is a difference of noisy ‘internal responses’ to the two stimuli: $r = r_1 - r_2$. We assumed there are three independent sources of noise in the task. First, there is noise related to the coding of the features, which corrupts the responses to both the target and the probe stimuli. Second, we assumed two sources of zero-mean noise related to memory: noise related to the delay or retention period itself, and noise that increases with memory load. Convolution of two wrapped Cauchy distributions is again a wrapped Cauchy distribution with a concentration parameter that is the product of the concentration parameters of the original distributions⁴⁷. The effect of memory noise can thus be modeled as:

$$\rho = \rho_{\text{feature}}^2 \rho_{\text{delay}} \rho_{\text{load}}^n \quad (2)$$

where ρ_{feature} reflects the precision in the representation of the eye or mouth region (ρ_{eyes} or ρ_{mouth} , squared because on each trial there are two stimuli, the target and the probe), ρ_{delay} is the precision of the noise due to the retention period, and ρ_{load} is for the noise due to each of the n additional items kept in memory. In the perception condition, ρ_{delay} and ρ_{load} are both equal to 1. We fit this model to the precision estimates extracted from the data from isolated eyes and mouth conditions to estimate the noise related to the two features, memory delay, and memory load, separately for each observer. We set the mean to zero (location parameter $\mu = 0$) for all distributions.

We then predicted the observer’s performance in the whole face task assuming optimal integration of the features (i.e. maximum likelihood estimation of identity given responses to features). As we did not have ‘cue-conflict’ conditions, that is, the eye and mouth region always had the same identity, we assume the responses to the two features have the same mean. To model the perception condition, we took the precision estimates for the features ρ_{eyes} and ρ_{mouth} and simulated 10^4 ‘trials’ by drawing random samples (‘responses’) from the corresponding wrapped Cauchy distributions. For each response, we computed the likelihood function for the true stimulus value given the response. As we assumed independent processing of the features, the combined likelihood is simply the product of the feature-specific likelihoods:

$$p(r_{\text{eyes}}, r_{\text{mouth}} | I) = p(r_{\text{eyes}} | I) p(r_{\text{mouth}} | I) \quad (3)$$

where I is the stimulus value (identity). The model observer made a maximum likelihood estimate of the identity by picking the value of I that maximized this likelihood. To quantify the model observer’s performance, we fit a wrapped Cauchy distribution to the errors in the same way as we did for the human data.

To model the memory conditions, we devised two versions of the optimal integration model, which differ in how memory noise corrupts the responses: In the first model, the features are first integrated, and the maximum likelihood estimate is then corrupted by memory noise. In the second model, the responses to the features are first corrupted by memory noise before being integrated. Additionally, we tested a model with equal weighting of the two cues (as we did not systematically manipulate cue reliability, this is likely to be very close to the optimal-integration model), and a model where the observer chooses the most reliable cue.

Data Availability

The data is available in Open Science Framework repository (osf.io/v79h6/).

References

1. Crouzet, S. M., Kirchner, H. & Thorpe, S. J. Fast saccades toward faces: face detection in just 100 ms. *J Vis* **10**(16), 11–17, <https://doi.org/10.1167/10.4.16> (2010).
2. Duchaine, B. & Yovel, G. A Revised Neural Framework for Face Processing. *Annual Review of Vision Science* **1**, 393–416, <https://doi.org/10.1146/annurev-vision-082114-035518> (2015).
3. Tsao, D. Y. & Livingstone, M. S. Mechanisms of face perception. *Annual review of neuroscience* **31**, 411–437, <https://doi.org/10.1146/annurev-neuro.30.051606.094238> (2008).
4. Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends in cognitive sciences* **4**, 223–233 (2000).

5. Leopold, D. A., Bondar, I. V. & Giese, M. A. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* **442**, 572–575, <https://doi.org/10.1038/nature04951> (2006).
6. Carlin, J. D. & Kriegeskorte, N. Adjudicating between face-coding models with individual-face fMRI responses. *PLoS computational biology* **13**, e1005604, <https://doi.org/10.1371/journal.pcbi.1005604> (2017).
7. Chang, L. & Tsao, D. Y. The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013–1028 e1014, <https://doi.org/10.1016/j.cell.2017.05.011> (2017).
8. Sekuler, A. B., Gaspar, C. M., Gold, J. M. & Bennett, P. J. Inversion leads to quantitative, not qualitative, changes in face processing. *Curr Biol* **14**, 391–396, <https://doi.org/10.1016/j.cub.2004.02.028> (2004).
9. Schyns, P. G., Bonnar, L. & Gosselin, F. Show me the features! Understanding recognition from the use of visual information. *Psychological science* **13**, 402–409, <https://doi.org/10.1111/1467-9280.00472> (2002).
10. Logan, A. J., Gordon, G. E. & Loffler, G. Contributions of individual face features to face discrimination. *Vision Res* **137**, 29–39, <https://doi.org/10.1016/j.visres.2017.05.011> (2017).
11. Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. N. What is “special” about face perception? *Psychol Rev* **105**, 482–498 (1998).
12. Richler, J. J. & Gauthier, I. A meta-analysis and review of holistic face processing. *Psychol Bull* **140**, 1281–1302, <https://doi.org/10.1037/a0037004> (2014).
13. Tanaka, J. W. & Farah, M. J. Parts and wholes in face recognition. *Q J Exp Psychol A* **46**, 225–245 (1993).
14. Taubert, J., Apthorp, D., Aagten-Murphy, D. & Alais, D. The role of holistic processing in face perception: evidence from the face inversion effect. *Vision Res* **51**, 1273–1278, <https://doi.org/10.1016/j.visres.2011.04.002> (2011).
15. Gold, J. M. *et al.* The perception of a familiar face is no more than the sum of its parts. *Psychonomic bulletin & review* **21**, 1465–1472, <https://doi.org/10.3758/s13423-014-0632-3> (2014).
16. Gold, J. M., Mundy, P. J. & Tjan, B. S. The perception of a face is no more than the sum of its parts. *Psychological science* **23**, 427–434, <https://doi.org/10.1177/0956797611427407> (2012).
17. Shen, J. & Palmeri, T. J. The perception of a face can be greater than the sum of its parts. *Psychonomic bulletin & review* **22**, 710–716, <https://doi.org/10.3758/s13423-014-0726-y> (2015).
18. Dobs, K., Ma, W. J. & Reddy, L. Near-optimal integration of facial form and motion. *Sci Rep* **7**, 11002, <https://doi.org/10.1038/s41598-017-10885-y> (2017).
19. Fougnie, D., Suchow, J. W. & Alvarez, G. A. Variability in the quality of visual working memory. *Nature communications* **3**, 1229, <https://doi.org/10.1038/ncomms2237> (2012).
20. van den Berg, R., Shin, H., Chou, W. C., George, R. & Ma, W. J. Variability in encoding precision accounts for visual short-term memory limitations. *Proc Natl Acad Sci USA* **109**, 8780–8785, <https://doi.org/10.1073/pnas.1117465109> (2012).
21. Bays, P. M. & Husain, M. Dynamic shifts of limited working memory resources in human vision. *Science* **321**, 851–854, <https://doi.org/10.1126/science.1158023> (2008).
22. Wilken, P. & Ma, W. J. A detection theory account of change detection. *J Vis* **4**, 1120–1135 (2004). 10:1167/4.12.11.
23. Zhang, W. & Luck, S. J. Discrete fixed-resolution representations in visual working memory. *Nature* **453**, 233–235, <https://doi.org/10.1038/nature06860> (2008).
24. Wheeler, M. E. & Treisman, A. M. Binding in short-term visual memory. *Journal of experimental psychology. General* **131**, 48–64 (2002).
25. Bays, P. M., Wu, E. Y. & Husain, M. Storage and binding of object features in visual working memory. *Neuropsychologia* **49**, 1622–1631, <https://doi.org/10.1016/j.neuropsychologia.2010.12.023> (2011).
26. Fougnie, D., Asplund, C. L. & Marois, R. What are the units of storage in visual working memory? *J Vis* **10**, 27, <https://doi.org/10.1167/10.12.27> (2010).
27. Shin, H. & Ma, W. J. Visual short-term memory for oriented, colored objects. *J Vis* **17**, 12, <https://doi.org/10.1167/17.9.12> (2017).
28. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Real-world objects are not represented as bound units: independent forgetting of different object details from visual memory. *Journal of experimental psychology. General* **142**, 791–808, <https://doi.org/10.1037/a0029649> (2013).
29. Curby, K. M. & Gauthier, I. A visual short-term memory advantage for faces. *Psychonomic bulletin & review* **14**, 620–628 (2007).
30. Lorenc, E. S., Pratte, M. S., Angeloni, C. F. & Tong, F. Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, perception & psychophysics*. <https://doi.org/10.3758/s13414-014-0653-z> (2014).
31. van den Berg, R., Awh, E. & Ma, W. J. Factorial comparison of working memory models. *Psychol Rev* **121**, 124–149, <https://doi.org/10.1037/a0035234> (2014).
32. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nat Neurosci* **17**, 347–356, <https://doi.org/10.1038/nn.3655> (2014).
33. Bays, P. M. Noise in neural populations accounts for errors in working memory. *J Neurosci* **34**, 3632–3645, <https://doi.org/10.1523/JNEUROSCI.3204-13.2014> (2014).
34. Parra, M. A., Della Sala, S., Logie, R. H. & Morcom, A. M. Neural correlates of shape-color binding in visual working memory. *Neuropsychologia* **52**, 27–36, <https://doi.org/10.1016/j.neuropsychologia.2013.09.036> (2014).
35. Pertzov, Y. & Husain, M. The privileged role of location in visual working memory. *Attention, perception & psychophysics* **76**, 1914–1924, <https://doi.org/10.3758/s13414-013-0541-y> (2014).
36. Schneegans, S. & Bays, P. M. Neural Architecture for Feature Binding in Visual Working Memory. *J Neurosci* **37**, 3913–3925, <https://doi.org/10.1523/JNEUROSCI.3493-16.2017> (2017).
37. Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science* **15**, 106–111, <https://doi.org/10.1111/j.0963-7214.2004.01502006.x> (2004).
38. Jiang, Y. V., Shim, W. M. & Makovski, T. Visual working memory for line orientations and face identities. *Perception & psychophysics* **70**, 1581–1591, <https://doi.org/10.3758/PP.70.8.1581> (2008).
39. Young, A. W. & Burton, A. M. Are We Face Experts? *Trends in cognitive sciences* **22**, 100–110, <https://doi.org/10.1016/j.tics.2017.11.007> (2018).
40. Langner, O. *et al.* Presentation and validation of the Radboud Faces Database. *Cognition and Emotion* **24**, 1377–1388, <https://doi.org/10.1080/02699930903485076> (2010).
41. Ebner, N. C., Riediger, M. & Lindenberger, U. FACES—a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav Res Methods* **42**, 351–362, <https://doi.org/10.3758/BRM.42.1.351> (2010).
42. Brainard, D. H. The Psychophysics Toolbox. *Spatial vision* **10**, 433–436 (1997).
43. Fisher, N. I. *Statistical Analysis of Circular Data*. (Cambridge University Press 1995).
44. Pewsey, A. The wrapped stable family of distributions as a flexible model for circular data. *Computational Statistics & Data Analysis* **52**, 1516–1523, <https://doi.org/10.1016/j.csda.2007.04.017> (2008).
45. JASP team (Version 0.9.0.1) (2018).
46. Wagenmakers, E. J. *et al.* Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic bulletin & review* **25**, 58–76, <https://doi.org/10.3758/s13423-017-1323-7> (2018).
47. Mardia, K. V. & Jupp, P. E. Directional statistics. Vol. 494 (John Wiley & Sons 2009).

Acknowledgements

This work was supported by the Academy of Finland (grant number 298329).

Author Contributions

K.Ö. T.S. and V.S. designed the experiments, analyzed the data and wrote the manuscript. I.M. prepared the stimuli, I.M. and K.Ö. conducted the measurements. V.S. and T.S. prepared the figures. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019