Vision Research 194 (2022) 107959

Contents lists available at ScienceDirect

Vision Research

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

The sizable difficulty in matching unfamiliar faces differing only moderately in orientation in depth is a function of image dissimilarity

Catrina M. Hacker[a], Irving Biederman[a,b,*], Tianyi Zhu[b], Miles Nelken[a], Emily X. Meschke[a]

[a] Program in Neuroscience, University of Southern California, USA
[b] Department of Psychology, University of Southern California, USA

ARTICLE INFO

Keywords:
Face recognition
Matching depth-rotated faces
Gabor-jet model
Face similarity
Match-to-sample

ABSTRACT

Attempting to match unfamiliar, highly similar faces at moderate differences in orientation in depth is surprisingly difficult. No neurocomputational account of these costs that addressed the representation of faces by which a face-similarity metric can be derived has been offered. A metric specifying the similarity of the to-be-distinguished faces is required as the rotation costs will be a function of the difficulty in distinguishing the faces. Consequently, rotation costs have typically been described in terms of angle of disparity, rather than the dissimilarity of the faces produced by the rotation. We assessed the effects of orientation disparity in a match-to-sample paradigm of a simultaneous presentation of a triangular display of three faces. Two lower test faces, a matching face and a foil, were always at the same orientation and differed by 0° to 20° from the sample on top. The similarity of the images was scaled by a model based on simple cell tuning, modeled as Gabor wavelets, that correlates almost perfectly with psychophysical similarity. Two measures of face similarity, with approximately additive effects on reaction times, accounted for matching performance: a) the decrease in similarity between the images of the matching and sample faces produced by increases in their orientation disparity, and b) the similarity between the matching face and the selection of a particular foil. The 20° orientation disparity was sufficient to yield a sizeable 301 msec increase in reaction time. An implication of the results is that the activity in V1 produced by viewing a face is fed forward to areas responsible for the individuation of that face.

1. Introduction

One remarkable feature of face recognition is that familiar faces can be recognized from various viewpoints despite the large distortions of the 2D retinal projections of the face produced by the variations in orientation. Conversely, in the absence of salient distinguishing local features, recognition of similar, unfamiliar faces has consistently been shown to incur large costs when recognition or matching has to be achieved even over moderately different orientations in depth (Duchaine & Nakayama, 2006; Bruce et al., 1999; Biederman & Kalocsai, 1997; Hill et al., 1997; Troje & Bülthoff, 1996; Valentin et al., 1997; Hancock et al., 2000; Barense et al., 2010; Natu & O'Toole, 2015). Somewhat surprisingly, there has been little neurocomputational explanation as to why the disparity in the orientation of faces produces such sizeable costs.

The present study employed a minimal match-to-sample task in which subjects viewed a triangular display of three computer-generated faces (Fig. 1) and attempted to select which one of two lower test faces was identical in identity to the upper face. On some trials, the faces could all be at the same orientation, in which case the image of the matching test face was identical to the sample. On other trials, the test faces differed in orientation in depth from the sample which meant that the images of the sample and the correct test face differed, although the identity was the same. The test face differing in identity from the sample served as the foil and could vary in similarity to the matching face but was always at the identical orientation to the matching face.

Some studies of the costs of orientation disparity on face recognition have used multiple foils (e.g., Duchaine & Nakayama, 2006; Bruce et al., 1999) typically held in memory, rendering it difficult, if not impossible to isolate an effect of distractor similarity. By employing only a single distractor which was in view during the matching, the present paradigm allowed a quantitative assessment of distractor similarity on the matching of face *percepts* rather than the memory of those percepts. The matching of faces at different disparities in depth could be separated into two quantitative measures of similarity between pairs of faces: a) The dissimilarity between the sample and matching test face produced by

[*] Corresponding author at: Department of Psychology, University of Southern California, USA.
E-mail address: bieder@usc.edu (I. Biederman).

https://doi.org/10.1016/j.visres.2021.09.005
Received 7 July 2020; Received in revised form 20 September 2021; Accepted 24 September 2021
Available online 16 February 2022

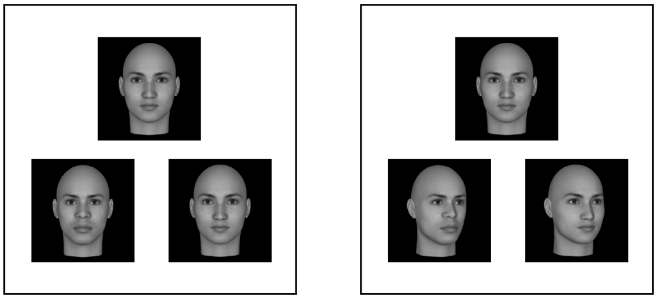

Fig. 1. Sample displays from two different trials with identical sample and test faces. Left panel: An example of the match-to-sample task with the sample (top) and the two test faces (bottom) presented at the same frontal (0°) orientation. Right panel: An example of a 20° trial in which the same test faces from the left panel differ from the same (0°) sample face (top) by 20°. The reader may sense the increased difficulty in matching when the sample and test faces are at disparate orientations in depth as the identities are the same in both panels. Both 0° and rotated trials (either 13° or 20°) could have the sample at any of the three orientations (0°, 13°, or 20°). The left–right designation of the stimuli refers to the *viewer's* (rather than the head's) left–right orientation. The normalized Gabor dissimilarity (Margalit et al., 2016) between the matching and foil faces was 2.74 for the left panel and 2.93 for the right panel. The Gabor dissimilarity between the matching test face and the sample is 0 in the left panel and 5.57 in the right panel. In both cases the correct match to the sample is on the right and the foil is on the left.

orientation disparity, and b) the similarity of the foil to the matching test face, dependent on the particular face selected as a foil on that particular trial, which determined the discrimination challenge. The two test faces—matching and foil—were always at the same orientation in depth which might or might not match the orientation of the sample. The greater the dissimilarity of the sample to the matching face produced by the rotation and the smaller the dissimilarity between matching face and foil based on the selection of a foil, the greater the expected difficulty in selecting the correct test face. A quantitative scaling of these two variables had, heretofore, never been evaluated, either individually or in concert.

The Gabor-jet model (Lades et al., 1993; Margalit et al., 2016) provides a means for scaling the similarity of images of faces based on a model of V1 simple cell filtering. The speed and accuracy of matching faces that are all at the same orientation is almost perfectly predicted by the similarity values of the model, with correlations with error rates in the mid 0.90s, even without a correction for the unreliability of the behavioral data (Yue et al., 2012). Although the Gabor-jet model of face dissimilarity is based on the multiscale, multiorientation of V1 simple cell coding, its exceptional predictability of the psychophysics of face discrimination suggests that the Gabor coding of faces in V1 is fed forward to face selective areas, such as FFA and OFA, that are critical for the individuation of faces. Additional justification for the Gabor-jet scaling of the similarity of faces derives from Yue et al. (2006) who showed that the representation of faces in FFA, a cortical area critical for individuating faces (Kanwisher & Yovel, 2006; Grill-Spector et al., 2004), is highly sensitive to the specific spatial (Fourier) kernels specifying the orientation, scale, and position of contrast that distinguish one image of a face from another. This sensitivity to the specific spatial content was not evident in the matching of complex blobs resembling teeth (Yue et al., 2006). It is the *image* similarity rather than the extraction of the underlying 3D representation of a face that is relevant to psychophysical matching. Thus, the advantage of matching bilaterally symmetrical faces is reduced when the faces are illuminated by asymmetrical lighting (Troje & Bülthoff, 1998).

The design of the present study employed a scaling of the dissimilarity between the matching test face and the sample as well as the matching test face and the foil (which were always at the same orientation). This allowed a test of whether the model's similarity values would also be highly predictive of performance when the faces were at different orientations in depth. In the absence of a principled measure of face similarity, past attempts at explanations of face rotation costs typically interpreted the costs in terms of viewing angle per se. This implicitly assumes a "protractor-in-the-head" representation in which matching is achieved through mental rotation or an alignment of a subset of features of one image to a subset of features of another stimulus. However, with a quantitative measure of face similarity one can a) assess the extent to which matching speed and accuracy is a positive function of the overall similarity of the matching test face to the sample and b) a negative function of the similarity of the foil to the matching stimulus, without a commitment to a particular angular transformation (which is difficult to implement without distinguishing local features to serve as landmarks). The investigation assessed the extent to which these two measures of image dissimilarity could account for human performance in matching faces differing in orientation. Because the faces were in view as the participants were trying to distinguish match from foil faces, the task assessed *perceptual* rather than *memorial* processing. (Simultaneous presentations were also used in the Hancock et al. and the Barense et al. studies cited above.)

One advantage of the match-to-sample paradigm over the oft used same-different judgment task is that subjects do not have to adopt an arbitrary criterion as to whether two highly similar faces are identical or not. Rather, a relative criterion—Which face more closely resembles the sample? —suffices. This criterion can be adopted because, unlike memory tasks with more than one possible face, the foil is well defined and in view so the similarity of the foil to the matching face can be calculated and its effects on performance evaluated. By using computer generated faces, the presence of local, distinguishing features that are abundant in photographs, such as a beauty mark, blemish, or the configuration of eyebrow hairs, could be excluded. The absence of such features meant that face matching in the present investigation required perception and discrimination of subtle metrically-varying configurations of attributes, such as the height of the cheekbones and the precise positions and shapes of the eyes, nose, and mouth.

2. Methods

2.1. Stimuli

The faces were created using FaceGen Modeller (Singular Inversions, Toronto, Canada), a 3D face modeling program. The core image was that of a bald, 20-year-old, Caucasian, gender neutral individual on a black background (Fig. 1). Twenty different identities were generated by varying the distances between eyes, nose, and mouth; height/prominence of cheek bones; jaw width; and very slightly varying the length and width of face parts such as the eyes, nose, and mouth (Fig. 2). The faces could be rotated 0°, 13°, or 20° in depth. The variations were metric, such as the degree of curvature of the eyebrows, rather than qualitative (or nonaccidental), such as whether eyebrows were curved or straight. These subtle variations of the default face were made to render the differences between the faces largely ineffable as occurs with naturally similar faces (Biederman & Kalocsai, 1997). The computer generation of the faces excluded the presence of local distinguishing features, such as beauty marks or moles, which would have allowed the subjects to zoom in and employ such features for distinguishing the faces rather than processing the whole face. The generation of the faces also avoided noticeable differences in standard population-defined categories (PDC) such as sex, race, age, expression or attractiveness which could have been employed to select a response. All the stimuli were in grey scale and were 256 × 256 pixels in extent.

Of a possible 190 combinations of matching-foil pairs of faces (disregarding status as matching or foil), only the 180 combinations with a normalized dissimilarity value of 1.50 or greater were used. The Gabor values were normalized by dividing the net similarity values by the number of jets–100 employed in this investigation–which yields the

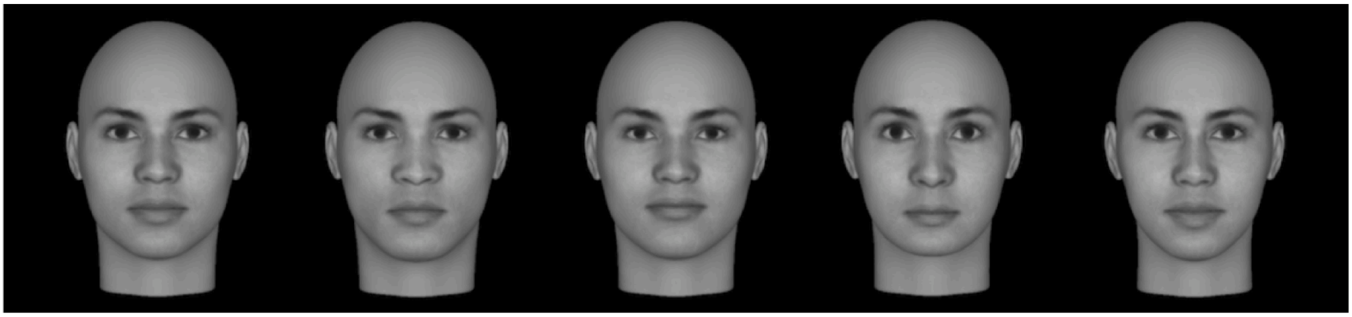

Fig. 2. Five examples of the 20 experimental faces. Subtle variation in the features created slightly different appearing individuals. The normalized Gabor dissimilarity[9] between faces 1 and 2 is 1.49 and between 1 and 5 is 3.63.

average jet dissimilarity value. The maximum dissimilarity value for pairs of faces in this experiment appeared to be approximately 5.5 normalized Gabor dissimilarity units. Each of these pairs of faces appeared twice throughout the experiment: once with the first face as the sample and a second time with the second face as the sample, for a total of 360 trials. The dissimilarity values between the matching and distractor faces were divided into four bins with an equal number of trials in each bin.

2.2. *Design and procedure*

Subjects performed the task which consisted of 360 2AFC match-to-sample trials on testable.org. Prior studies had established that the performance of online subjects was highly similar to those personally run in the lab except that an occasional online subject failed to exercise due diligence in performing the task. On each trial, subjects viewed a triangular arrangement of three faces with one face (the sample) centered above two lower faces (the test stimuli), one of which matched the identity of the sample (Fig. 1). The orientation in depth of the test faces could differ from the sample face by 0°, 13°, or 20°. The two test faces were always at the same orientation in depth. Because the orientation in depth of the test stimuli could differ from the sample, the image of the matching face could differ from the sample, but the identity was always an exact match. For trials where the sample and test faces differed in orientation, the departure from the 0° orientation could be implemented in the sample or the test faces, e.g., the sample could be at 0° and the two test faces at 13°, or the sample could be presented at 13° and the two faces could be at 0°. In either case, the test faces would be rotated 13° from the sample (an orientation *disparity* of 13°) and were so classified. Faces were always rotated to their left (subject's right) as in the right panel of Fig. 1. Subjects indicated which of the two test faces matched the identity of the sample by pressing the left or right arrow key as quickly and as accurately as possible. The orientation disparities were approximately balanced over the various levels of matching and foil similarity values.

The stimuli were displayed for 5 s, although responses were recorded after the cessation of the display on the rare occasion that response time exceeded 5 s and the next trial had not yet started. Reaction times that were shorter than 500 msec or longer than 10.0 s were not included in the data analysis. Within and across each block, the stimuli were balanced by face identity (one of the 20 faces that differed in underlying shape), Gabor dissimilarity between 180 combinations of matching and distractor faces, and orientation condition (e.g., 0°- 0°, 0°- 13°, 0°- 20°, 13°- 13°, etc.), for which there were 60 trials per condition (the 20°- 20° condition was not used). All subjects viewed the same stimuli presented in different random orders. Subjects could pause at their leisure between any of the five blocks, although not between individual trials within a block. The total time for testing was approximately 25 min, which included 5 min for instructions.

2.3. *Displays*

To ensure that all images were displayed within the boundaries of the screen independent of the particular computer used by a subject, a calibration procedure was run at the start of the experiment. Each subject used the left and right arrow keys to adjust the length of a line on the screen to match the horizontal extent of a standard credit card. Subjects were instructed to sit at a viewing distance of approximately an arm's length from the computer screen. At this distance from a 15″ laptop screen, each face was bounded by a square that subtended a visual angle of approximately 5.0° on each side with a horizontal separation of 0.7° between the lower two test headshots and a vertical separation of 0.7° between the test and sample headshots (Fig. 1). An important design feature of the display was the diagonal arrangement of the faces which defeats local pixel- or feature-based comparison processes which could be more readily engaged if the faces were aligned vertically or horizontally.

2.4. *Stimulus similarity scaling*

The Gabor-Jet scaling of the physical dissimilarity of pairs of faces was computed from a 10 × 10 grid centered on each face. The procedure is illustrated in Margalit et al. (2016) which presents an app for computing the Gabor dissimilarity of faces. Each node of the grid corresponds to the center of the receptive fields of the kernels of one jet (modeling the orientation and scale tuning of a single, simplified V1 hypercolumn) composed of 80 Gabor filters at 8 equally spaced orientations (22.5° differences in angle), 5 scales, and 2 phases (sine and cosine). The coefficients of the kernels (with the magnitude representing the activation value of a single simple cell) within each jet were then concatenated to an 8000-element vector representing each image (100 jets × 80 kernels). Image similarity was computed as the Euclidean distance between two 8000-value vectors. When images of two faces are identical, their dissimilarity is zero.

2.4.1. *Origin of the parameters in the Gabor-jet model*

The selection of 100 jets followed the von der Malsburg's group finding that ceiling levels of recognition accuracy could be reached with 100 jets (often less) and that eight orientations and five scales were similarly sufficient and were compatible with psychophysical data from H. R. Wilson and associates (Wilson & Bergen, 1979; Wilson et al., 1983; Phillips & Wilson, 1984). Five to six spatial frequency channels and approximately 8 orientation channels appeared to capture all the data, while more would be redundant. These spatial frequency and orientation parameters agreed with single unit recordings in monkey V1 (De Valois & De Valois, 1980).

It should be noted that there is considerable overlap of the receptive fields of the cells (kernels) of one jet with the kernels of nearby jets as well as the cells with large r.f.s which cover much of the face. There is overlap with multiple jets distributed all over the face so any one region of the face is multiply encoded by different jets which can be the basis of face configural effects (Xu et al., 2014).

As noted previously, the similarity values calculated in this manner predict human psychophysical similarity in a match-to-sample task for faces at the same orientation in depth almost perfectly (Yue et al., 2012), with correlations between Gabor dissimilarity values and error rates of approximately r = −0.95, without allowance for the unreliability of the participants' data. The Gabor-jet dissimilarity values in the present experiment between matching and foil faces on individual trials at 0° orientation disparity were selected to fall within a range of 1.50 to 4.12 which, for most subjects, placed them in an intermediate range of difficulty that allowed performance to reflect the experimental variations. Yue et al. (2012) showed that this range produced error rates from near chance ~ 45% (chance = 50%) to ~ 1% in the match-to-sample task with faces at the same orientation. This was generally true of the present study. *The Gabor-jet model thus describes a nearly perfect lawful relationship, between a physical measure of highly complex, ineffable stimuli (the Gabor coding of the similarity of faces) and a psychological measure (the difficulty of distinguishing those faces)*. Prior examples of high correlations between psychophysical judgments and physical dimensions, such as those documenting Weber's Law, were with unidimensional physical dimensions, such as loudness or weight.

We note that the dissimilarity values reflect only the physical differences between a pair of faces as registered by the Gabor kernels and are not sensitive, by themselves, to *population-defined categories* (PDCs) such as sex, age, expression, race, or attractiveness which can be learned from exposures to samples of faces. Using reverse correlation, a classifier for each of these population-defined categories can be calculated from the pattern of activation of the various Gabor-like kernels for these categories (Mangini & Biederman, 2004). Prosopagnosics can typically identify the PDCs of a face, e.g., that it is of a 20ish year-old Caucasian female, but they fail at individuating a face within its PDC. An acquired prosopagnosic with severe bilateral lesions to cortical face areas OFA and FFA is unable to recognize his wife, siblings, or his own face in a mirror but is normal in distinguishing and describing population-defined attributes of faces (Mangini & Biederman, 2004; Xu & Biederman, 2014). This result suggests that the recognition of PDCs engage different networks than those involved in individuating a face and it is face individuation that is the subject of the present investigation. However, in noting the preservation of PDC perception in individuals who are unable to individuate faces, the possibility remains that prosopagnosics may terminate the difficult processing required for face individuation while able to determine the more readily available features signaling PDCs.

The dissimilarities of two kinds of relations between the faces on each trial are of particular relevance: a) the dissimilarity between the matching (correct) test face and the sample, reflecting the difference in orientation, and b) the dissimilarity between the matching and foil test faces. The average dissimilarities over orientation angle are shown in Fig. 3. The dissimilarity between the matching and sample faces increased markedly with increasing orientation disparities, with most of the increase occurring between 0° and 13° with a somewhat smaller increase in dissimilarity between 13° and 20°. We also considered an alternative scaling method, the Fiducial Point Model (FPM) (Wiskott et al., 1997; Müeller & Wuertz, 2009), in which the individual jets are not centered in a regular grid, with a somewhat arbitrary positioning of jets with respect to face features, but are positioned over specific facial landmarks (termed *fiducial points*), such as the pupil of the left eye or the tip of the nose. The FPM yielded a similar pattern of dissimilarities as those shown in Fig. 3 using the grid model with some puzzling exceptions in that for some faces there was not a monotonic increase in matching-sample dissimilarity with increasing orientation disparities. In general, the FPM did not provide as good a qualitative fit to the data as the grid model in that it failed to show greater costs with increasing orientation disparities and it failed to reflect the slightly greater similarity of the matching (vs. the foil) face to the sample under rotation. Consequently, all analyses reported here employed the grid model. Later we consider why the explicit locations of facial landmarks encoded in the FPM did not provide as good a match to the data as the grid model.

A straightforward expectation would be that the difficulty in matching faces (i.e., longer RTs and higher error rates) would increase with increasing dissimilarity between the matching face and the sample. We would also expect that an increase in similarity between matching and foil stimuli would render matching more difficult, resulting in longer RTs and higher error rates. Fig. 3 also shows that the sizable

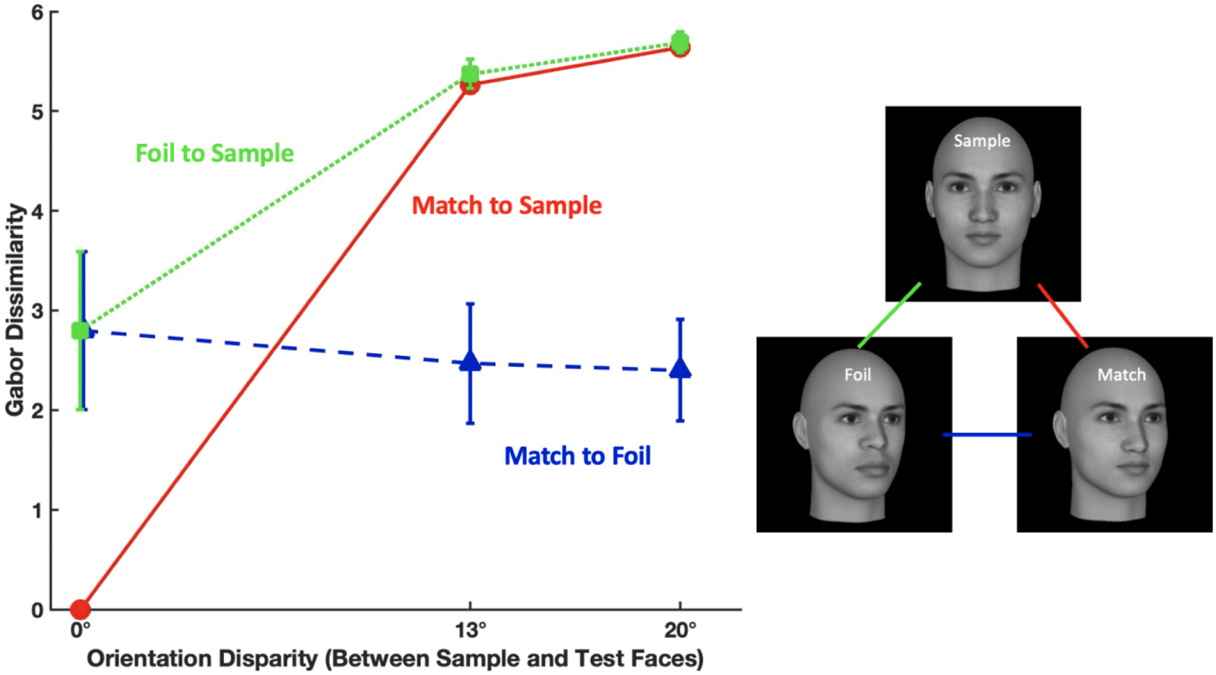

Fig. 3. Mean normalized Gabor dissimilarity values for the three stimulus relations on each trial (Match to Sample, Foil to Sample, and Match to Foil) as a function of the orientation difference between the sample and test faces. The matching and foil test faces were always presented at the same orientation. The error bars are the standard deviations of the mean of the Gabor dissimilarity values at each orientation difference between the various instances of the 20 sample and test faces.

difference in dissimilarity of about 3.00 Gabor units between matching and foil test faces to sample and foil to sample dissimilarities at 0° disparity (reflecting that the matching face is identical to the sample whereas the selection of the face that serves as a foil differs by an average of 3.00 Gabor units from the sample), virtually disappears at orientation disparities between sample and test faces at 13° and 20°. This effect is likely a consequence of the shifting of surfaces produced by even modest rotations in depth (which maxed out at 13° in the present experiment) so that a given surface is no longer closely aligned in the grid of Gabor jets. Such a shift could override the subtle changes distinguishing target and matching faces. If the matching of the correct test face to the sample is based on Gabor similarity, the loss of the characteristics that distinguished the matching from the foil faces at the modest orientation disparities suggests that there should be a marked increase in difficulty in judging which of the two test faces matches the sample.

As shown in Fig. 3, unlike the dissimilarity of matching to sample faces, the average dissimilarity between the 20 matching and foil test faces remained relatively constant, except for some slight diminution in the variance over rotation angle between sample and test faces. The matching and foil faces were always presented at the same orientation. One possibility for the reduced variance is that more of the identical cheek/ear region of the faces was being compared when both faces were at 20° than at 0°. This would account for the slight decrease in overall dissimilarity as well.

Nonetheless, within a given orientation there was considerable variation in dissimilarity values among the 180 combinations of different foil and matching faces as reflected in the error bars. This variation allowed a straightforward test of whether the effect of the dissimilarities of the foil to matching faces would be independent of the disparity in orientation between matching and sample faces.

2.5. Participants

Sixty-five participants (mean age 20.55 yrs., range 18–47 years, 16 males) performed the web-based USC Rotated Face Perception Test, (USC rFPT), for course credit in the Department of Psychology subject pool. Six of these subjects were excluded from further analysis: five had more than 20 correct trials with a reaction time below 750 ms, and one had more than 5 trials with a reaction time greater than 7.5 s. Subjects excluded based on fast reaction times (<750 msec) all had extremely high error rates (over 40%), indicative of a lack of engagement in the discriminative challenge. All subjects reported normal or corrected-to-normal vision and no neurological or visual disorders. The work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All subjects gave informed consent in accordance with the procedures approved by University of Southern California's University Park Institutional Review Board.

3. Results

3.1. Effect of orientation disparity

Fig. 4 shows the mean correct RTs as a function of the Match-to-Foil Dissimilarity values separately for the three angular disparities (0°, 13°, or 20°) between sample and matching stimuli. For a given orientation disparity, the data are collapsed over the particular orientation disparities between the sample and the matching test faces. Thus, the data for when the sample face was at 0° and the matching face at 13° are combined with the data for when the sample face was at 13° and the matching face at 0° as there were only small and inconsistent differences in performance when the sample or test faces were rotated. When matching faces at different orientations, subjects appeared more willing

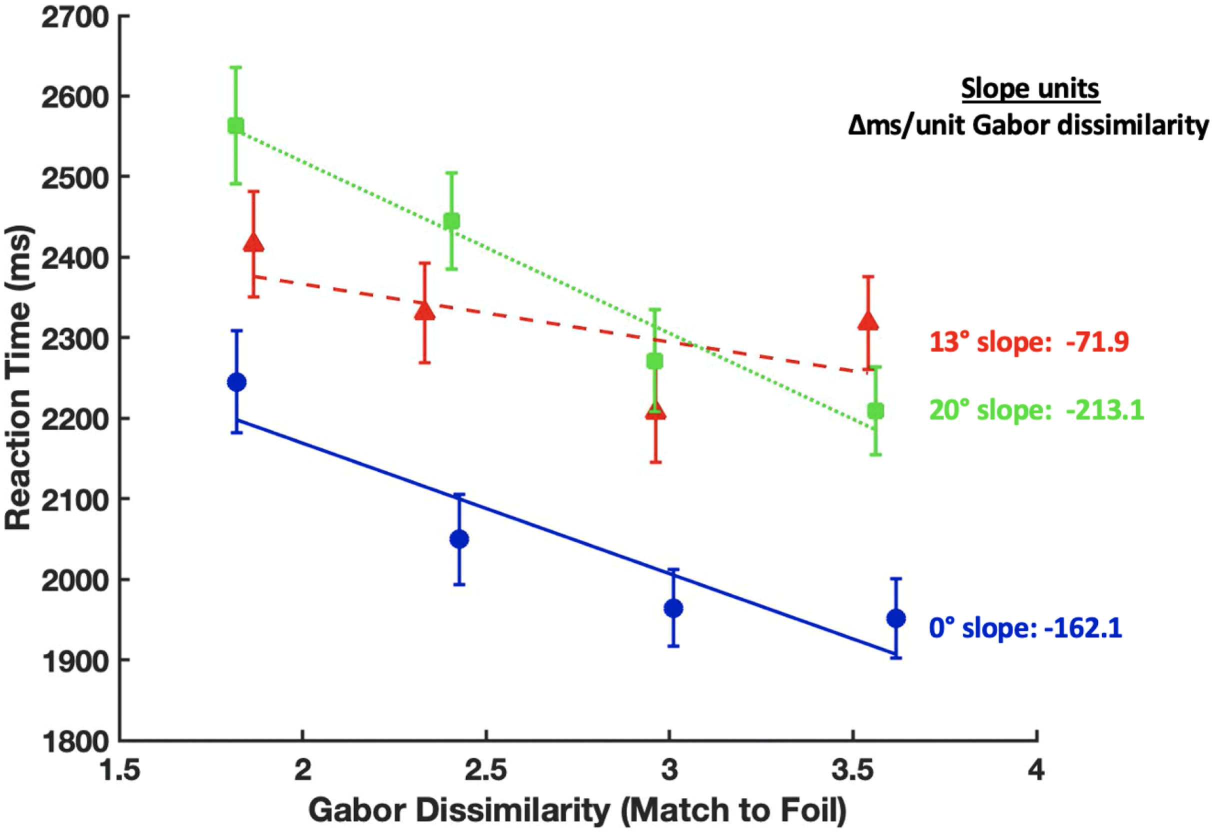

Fig. 4. Mean correct reaction time (msec) as a function of the normalized binned Gabor dissimilarity values between the matching and the foil faces over the three levels of orientation disparity. The slopes are in units of msec per unit of normalized Gabor dissimilarity. 0°: Solid line, round points; 13°: Dashed line, triangular points; 20°: Dotted line, square points.

to tolerate longer RTs than a higher error rate. An orientation disparity of 20° produced an increase in error rates of only 3.8% above that at an orientation disparity of 0° but a sizeable 301 msec increase in RTs. The greater the orientation difference between matching and sample faces, the greater the difficulty in matching as reflected in longer RTs and higher error rates. For orientation disparities between matching and sample faces of 0°, 13°, and 20°, mean RTs (and percent errors) were 2080 msec (14.5%), 2318 msec (17.2%), and 2381 msec (18.3%), for reaction times $F(2, 116) = 50.06$, $p < .001$, $\eta_p^2 = 0.463$, for error rates, $F(2,62) = 3.56$, $p < .05$, $\eta_p^2 = 0.236$. (η_p^2 = partial eta squared, a measure of the proportion of the total variance associated with an independent variable with the effects of other independent variables and interactions partialed out (Richardson, 2011)).

The response times in the present study were a bit over 2 s which would be longer than those in a two-choice task in which participants judged whether an image of a given face, say, was that of a familiar celebrity (Hacker et al., 2019). Such a decision could be readily accomplished with just a single fixation at a brief 100 msec presentation. However, our task required at least two or three "inspection" saccades to distinguish which of two highly similar faces differed from the sample (as can be appreciated by attempting the sample trials in Fig. 1) hence the need for relatively longer stimulus presentation times than those required for a judgment of whether a single face was that of a celebrity (Hacker et al., 2019). These inspection saccades tend to each be longer than the single fixations required to classify whether a face is familiar. The need for several inspection saccades likely contributed to the relatively long overall durations for the task and the time for these saccades elevated the overall mean time thus reducing the relative cost of the rotation itself.

3.2. *Match-to-foil dissimilarity effects*

The mean dissimilarity between matching and foil faces remained largely constant as orientation disparity increased, as shown in Fig. 3, so the cost of the disparities in orientation between sample and matching faces was wholly a function of the increased Gabor dissimilarity of the matching test face to the sample as orientation disparity increased. There was, nonetheless, considerable trial-to-trial variation in the match-to-foil similarities among the various pairs of the 20 matching and foil test faces within each orientation disparity. Data were grouped into four bins based on Gabor dissimilarity between matching and foil faces and mean correct RTs were computed for each bin at each orientation disparity as shown in Fig. 4. (These results were highly similar to those obtained with the raw data.) With the exception of the single aberrant point on the 13° orientation disparity function, the Match-to-Foil dissimilarity values were linearly related to RTs with greater dissimilarities (i.e. more easily distinguished faces) yielding shorter RTs, $F(3, 174) = 51.90$, $p < .001$, $\eta_p^2 = 0.472$. The point is termed "aberrant" as it had a mean RT that exceeded that of the two match-to-sample points on that same 13° function that had lower match-to-foil dissimilarity values and even exceeded the corresponding point on the 20° function which accounted for the shallow slope of the 13° function and the low correlation with match-to-foil dissimilarity relative to the 0° and 20° functions.

The Pearson correlations between the Match-to-Foil (bin) dissimilarity values and RTs were −0.92 for 0° disparity, −0.61 for 13° disparity and −0.98 for 20°. Correlations for unbinned data (not shown) showed similar trends to the binned data: $r = -0.44$ for 0° disparity, $p < 0.0001$, 95% CI = [−0.573, −0.284], $r = -0.15$ for 13° disparity, $p = 0.10$, 95% CI = [−0.32, 0.03], and r = -0.31 for 20° disparity, $p < 0.001$, 95% CI = [−0.46, −0.14]. The slope of the correlation between Gabor dissimilarity and RTs (in units of msec/unit of Gabor dissimilarity) was −0.160 for 0° disparity, −0.073 for 13° disparity and −0.158 for 20° disparity, showing the same trends as the binned data in Fig. 4.

The main effect of Orientation Disparity was highly significant, $F(2, 116) = 154.23$, $p < .001$, $\eta_p^2 = 0.727$, with the greater the dissimilarity between matching and sample faces (associated with greater angular disparity), the longer the RTs. Given the power in the analysis, the Orientation X Bin interaction was also significant, $F(6, 348) = 6.05$, $p < .001$ although the η_p^2 value was only 0.094, classified as a weak effect, attributable to the single point on the 13° orientation disparity function. The relative magnitude of the effect of the interaction can be appreciated in comparison to the η_p^2 values for the main effects of Orientation and Match-to-Foil (bin) dissimilarity which were 7.7 and 5.0 times, respectively, the magnitude of their interaction.

The most striking aspects of these data are that the reduced cost (decrease in RTs) is linear with an increase in Gabor dissimilarity between matching and foil faces. This effect is roughly independent of orientation disparity with the shallower slope of the 13° function attributable to the previously noted single aberrant point at the highest match-to-foil dissimilarity value being a departure from what otherwise would be additivity in the RTs for orientation disparity and Match-to-Foil dissimilarity. To assess the possible role of errors in this pattern of data, the dependent variable in Fig. 4 was plotted as inverse efficiencies in which each RT was divided by the accuracy (percent correct) at that point. The picture that emerged was virtually identical to the data in Fig. 4 (as shown in Supplementary Fig. 1) so differences in error rates are unlikely to be the cause of the shallower slope at 13°. This point is discrepant with the other data in the experiment and the results of prior studies consistently showing that the higher the match-to-foil Gabor dissimilarity, the shorter the RTs (Yue et al., 2012).

3.3. *Relative magnitude of the effects of variations in the match-to-sample and foil-to-sample similarities*

As noted above, performance (RTs) in this task was largely a function of two parameters: a) the similarity of the matching face to the sample, which would decrease with an increase in orientation disparity, and b) the similarity of the foil to the matching stimulus which would be unaffected by orientation disparity but would be a function of the particular faces selected for a given trial.

Compared to the 0° orientation difference between the matching stimulus and the sample, the rotation of 20° decreased the similarity between the matching stimulus and the sample by an average of 5.67 Gabor units (see Fig. 3) producing a 301 msec increase in RTs or an increase of 53 msec per unit of reduced Gabor similarity. As shown in Fig. 4, the mean slope of the three functions is −148 msec/unit of Gabor dissimilarity suggesting that the effect of similarity between the foil and matching stimulus (i.e., the discrimination challenge) is approximately 2.8 times the effect of dissimilarity between the matching test stimulus and the sample. Put another way, for each unit of decreased Gabor dissimilarity between matching and foil stimuli, the increase in RTs is 2.8 times greater than the increase in RTs produced by each unit of decreased Gabor similarity produced by the orientation disparity between sample and matching stimulus.

3.4. *Fiducial point scaling*

As noted earlier, an alternative scheme for the similarity scaling of faces is to use a Fiducial Point Model (FPM) in which each jet is centered on a particular face landmark, such as the pupil of the left eye or the tip of the nose. Although the FPM captures the observer's face knowledge as to corresponding face features in two faces at different orientations (e.g., placement of pupils of the eyes, tip of the nose), somewhat surprisingly, the FPM yielded *larger* Gabor dissimilarities compared to the grid model, without any noticeable gain in predictability. This is likely a consequence of the positioning of the jets in the fiducial model FPM at points of higher contrast variation than the jet locations in the grid model. The latter model has some jets centered on the black background or middle of the cheeks where the change in dissimilarity as the face was rotated in depth, for kernels with smaller receptive fields, would be very low. Of greater concern as to the adequacy of the FPM, for eight of the 20 faces, the FPM yielded *smaller* Gabor dissimilarities between 0° and 20° orientations of those faces compared to the dissimilarities between the 0° and 13°. In the grid model, all 20 faces were more dissimilar to the 0° face at 20°compared to 13°, in line with the data showing increased difficulty in matching over a 20° disparity than a 13° disparity.

Still another discrepancy of the FPM with the behavioral data was that at an orientation disparity of 13° it failed to reflect the (slightly) greater similarity of the matching face to the sample compared to the foil with the sample so it could not account for the increase in RTs.

It might seem surprising that the FPM in its explicit coding of aspects of face knowledge does not achieve any gain in predictability compared to the grid model. We propose that what the FPM renders explicit about the face, such as the locations of the pupils of the eyes or the corners of the mouth, are not what limits performance on the present task. We presume that all individuals with normal vision would be able to locate the various face landmarks. The challenge would be in recognizing the extent of the metric deformations in the images of faces when matching had to be executed at different disparities in orientations. The FPM's specification of the locations of facial landmarks would seem to have its greatest value in *finding* a face in an uncertain location rather than in determining its identification.

4. Discussion and conclusion

Although sizable costs in the perceptual matching of unfamiliar faces differing modestly in orientation have been reported in the literature, there had been no general explanation of these costs. We document that two stimulus parameters, both reflecting the physical similarity of pairs of faces as scaled by the Gabor-jet model, may be sufficient to provide an account of these costs: 1) The dissimilarity of the matching face to the sample produced by the orientation disparity between the two, and 2) the similarity of the foil to the matching face which defines the discrimination challenge. High similarity of the foil and matching faces increases the uncertainty as to which test face is a match to the sample. As scaled by the Gabor-jet model, increases in the values of each of these parameters produced increases in RTs. Moreover, the effects of the two were approximately additive on RTs (save for the single point on the 13° function). Insofar as the similarity of the foil to the matching face can be regarded as the perceptual challenge in this task, the approximate additivity with the effects of orientation disparity suggests, by additive factors logic (Sternberg, 1969; Sternberg, 2011) that the two factors affect different processing stages. We also note that the near perfect correlations (*r*s in the mid to high 0.90s without correction for unreliability) in prior experiments, e.g., Yue et al. (2012), with this match-to-sample paradigm with all faces at 0° suggests that no additional processing stages that embellish the representation of a face are required between V1 and the posterior face-selective areas such as FFA and OFA to account for the perceptual matching of faces by humans.

Some investigators (e.g., Swystun & Logan, 2019; Wilson et al., 2002) have employed synthetic faces of relatively low dimensionality that are designed to capture some of the natural variation between individual faces including PDCs. Studies with these faces show a number of the qualitative effects apparent with facial photographs and computer generated faces, such as rotation-in-depth costs, the Thatcher illusion, and inversion costs. Also, participants can accurately match the synthetic faces to the original photographs although at a coarser scale as is evident in the present study and in Yue et al. (2012). Whereas the Gabor-jet model was inspired by the Gabor coding in V1 hypercolumns (Barense et al., 2010; Biederman, 2000; Biederman & Bar, 1999; Lades et al., 1993) and Yue et al. (2006) showed that such coding was also true of FFA, it remains to be seen whether alternative schemes reflect neurocomputational processing in face selective areas in the cortex.

It is important to emphasize that the good quantitative fit of the Gabor jet model in accounting for the costs of orientation disparity and matching-foil similarity is dependent on the faces differing *only* in the metrics of their physical characteristics. This restriction excludes the faces differing in a) familiarity, b) PDCs such as sex, age, race, or attractiveness, and c) distinguishing nonaccidental properties (NAPs) such as the presence of a blemish or straight vs. curved eyebrows. The heightened sensitivity to differences in these three image variations arise from networks later in the ventral pathway than V1 (Ramon et al., 2015) and are thus not reflected in Gabor similarity. The exclusion of NAP differences between the faces means the faces must differ only in the metric properties of their parts and relations. When faces differ in NAPs the normal processes of face recognition in which a configural representation of the metric variations of faces is sufficient to elicit identity tends to be bypassed in favor of a search for a distinctive feature as is evident in the strategies of many prosopagnosics. That PDCs are appropriately to be distinguished from normal metric variation of faces is supported by the phenomenon, noted previously, that the deficit in prosopagnosia is largely confined to the individuation of a face, not in knowing its PDCs (Mangini & Biederman, 2004).

The present finding of sizeable costs when matching faces as a consequence of orientation disparity might be somewhat unexpected from a report that anterior face patches in the macaque show that pose-invariant face identity is maximal in those regions (Freiwald & Tsao, 2010). A possible resolution to this discrepancy is that the faces in the macaque study were photographs which likely differed in NAPs as well as PDCs. If the faces to be distinguished on a given trial in the present experiment would have differed in a NAP or a PDC, performance would likely have been close to errorless with very short reaction times.

4.1. The relation between the recognition of depth-rotated faces and the recognition of depth-rotated objects

The most significant determinant of the costs in matching or recognizing similar depth-rotated visual entities—faces or objects—is whether the stimuli to be distinguished differ in nonaccidental properties (Meyers et al., 2015; Biederman & Bar, 1999; Biederman, 2000). Nonaccidental properties (NAPs) are characteristics or features of a visual entity that are invariant with the orientation of the object in depth. Such properties can be an aspect of shape, such as whether a particular contour, say an eyebrow, is straight or curved, or they can be a characteristic of a surface feature of the object, such as whether there is or is not a dark spot, such as a blemish, in a given region of a face. If the stimuli can be distinguished by a NAP (or NAPs) and the object or face is at an orientation where the NAP is in view, then a rapid, correct response can be produced without even processing the face itself. NAPs are distinguished from metric properties (MPs) such as the degree of curvature of a contour or the angle of the junction between two cylinders. MPs vary continuously with orientation in depth. The Gabor jet measure is sensitive to MP variation in the image and does not reflect the markedly greater impact of NAP differences (treating them as metric differences) or to learned differences, such as that reflected in PDCs in distinguishing faces. People (and macaques) find it exceedingly difficult to individuate an object that differs from foils only metrically when the object is encountered at a new orientation in depth, as documented by the difficulty in matching wire frame objects resembling bent paper clips that were studied extensively in the 1980s and 1990s (Biederman, 2000; Logothetis et al., 1994). Cells in the inferior temporal region of the macaque modulate their firing much more to equal Gabor-jet image changes in a NAP, for example, from straight to curved, than changes in an MP, such as degree of curvature or differences in the angle of attachment of a pair of cylinders, suggesting a fundamental neural basis to the greater sensitivity to NAP compared to MP differences (Kayaert et al., 2003).

This phenomenon of greater nonaccidental than metric sensitivity in face recognition is underscored in the striking demonstrations of Sinha and Poggio (1996); Sinha and Poggio (2002) in which observers fail to notice the substitution of a President's inner face for the inner face of his vice president (Clinton and Gore; Bush and Cheney). In these instances, the inner facial features vary metrically and subtly between the president and his vice-president and are dominated by the larger qualitative, i.e., nonaccidental, differences in the external configuration of hairline and head shape as well as eyeglasses, and context.

The absence in our face stimuli of surface features or variations of face attributes that differed nonaccidentally were motivated by the desire to limit the variation between faces to metric properties. As noted previously, with pairs of faces that differ in local nonaccidental features, observers will focus on the region with the featural difference and thus circumvent normal face processing. Normal face processing is characterized by configural effects, produced by the overlap of large receptive fields (RFs) which are centered at different positions on the face. They allow the face to be processed as a whole with the RF overlap allowing multiple redundant coding of the small metric differences between similar faces (Xu et al., 2014). The present minimal match-to-sample paradigm can be employed to study the discrimination of any pair of stimuli, including metric variations of the harmonics of a sphere which produce smooth complex sculptured volumes resembling teeth. These have been employed as non-face comparison stimuli for studying what might be unique about faces. They show the same high correlations (in the mid -0.90 s) with RTs and error rates as the discrimination of faces.

4.2. Matching computer-generated vs. photos of faces

The present investigation employed computer generated (CG) images of faces. Crookes et al. (2015) showed that the "other race effect," a reduction in accuracy of face recognition when judging faces other than that of the participant's own race, was diminished with CG faces compared to photographs. It is possible that CG faces more generally show reduced sensitivity to PDCs, such as race, perhaps partly as a function of their reduced animacy (noted by Crookes et al.), but it would seem implausible that the discrimination of physical variations within a single population classification, as in the present paradigm, could be any more sensitive to the physical variations as assessed by the near ceiling correlations with Gabor similarity–without even any correction for the underlying unreliability of the behavioral data. This said, generalizations of results with CG faces need to be assessed with photographs despite the challenges of achieving control of local features and subtle differences in population-defined attributes.

4.3. Role of mental rotation?

We believe that our results are consistent with decisions made directly on the images to be discriminated based on their similarity, rather than employing mental rotation to align images at different orientations in depth. Humans and macaques find it exceedingly difficult to mentally rotate in depth complex images differing slightly in metric properties, such as the faces in the present experiment or the bent paper clip-like stimuli that were used in a number of experiments a few decades ago. Typically, the clips were five thin cylinders connected end to end which only differed in their angles of attachments. The inability of subjects to rotate such stimuli in depth led to the idea that object recognition was "view based", requiring exposure to different views of the clips that differed by more than a few degrees from the training stimuli. (NAP differences would negate the requirement of previous exposure to the different orientations.) This may be the reason why the high costs of a disparity in depth in distinguishing metrically varying faces only is apparent with unfamiliar faces.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2021.09.005.

References

Barense, M. D., Henson, R. N., Lee, A. C., & Grahm, K. S. (2010). Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: Effects of viewpoint. *Hippocampus, 20*, 389–401.

Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision, 13*(2-3), 241–253.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39*(17), 2885–2899.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philosophical Transactions of the Royal Society London: Biological Sciences, 352, 1203–1219.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339–360.

Crookes, K., Ewing, L., Gildenhuys, J.-D., Kloth, N., Hayward, W. G., Oxner, M., et al. (2015). How well do computer-generated faces tap face expertise? *PLoS One, 10*(11), e0141353. https://doi.org/10.1371/journal.pone.014135310.1371/journal.pone.0141353.g00110.1371/journal.pone.0141353.g00210.1371/journal.pone.0141353.g00310.1371/journal.pone.0141353.g00410.1371/journal.pone.0141353.g00510.1371/journal.pone.0141353.t00110.1371/journal.pone.0141353.t00210.1371/journal.pone.0141353.t00310.1371/journal.pone.0141353.t00410.1371/journal.pone.0141353.s001

De Valois, R. L., & De Valois, K. K. (1980). Spatial vision. *Annual Review of Psychology, 31*(1), 309–341.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*(4), 576–585.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science, 330*(6005), 845–851.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555–562.

Hacker, C. M., Meschke, E. X., & Biederman, I. (2019). A face in a (temporal) crowd. *Vision Research, 157*, 55–60.

Hancock, P. J. B., Bruce, V., & Burton, A. M. Recognition of unfamiliar faces. (2000). Trends in Cognitive Sciences, 4, 330–337.

Hill, H., Schyns, P., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201–222.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society London: Biological Sciences, B, 361*, 2109–2128.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferotemporal cortex. *Journal of Neuroscience, 23*, 3016–3027.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions of Computer Science, 42*(3), 300–311.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology, 4*(5), 401–414.

Margalit, E., Biederman, I., Herald, S. B., Yue, X., & von der Malsburg, C. (2016). An applet for the Gabor scaling of the differences between complex stimuli. *Attention, Perception, & Psychophysics, 78*, 2298–2306.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classification. *Cognitive Science, 28*, 209–226.

Meyers, E. M., Borzello, M., Freiwald, W. A., & Tsao, D. (2015). Intelligent information loss: The coding of facial identity, head pose, and non-face information in the macaque face patch system. *Journal of Neuroscience, 35*(18), 7069–7081.

Müeller, M. K. & Wuertz, R. P. Learning from Examples to Generalize over Pose and Illumination. In C. Alippi et al., Artificial Neural Networks, ICANN 2009, Part II, LNCS 5769, 643-652., Berlin Springer-Verlab. (2009).

Natu, V. S., & O'Toole, A. J. (2015). Spatiotemporal changes in neural response patterns to faces varying in visual familiarity. *Neuroimage, 108*, 151–159.

Phillips, G. C., & Wilson, H. R. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society of America, A, 1*(2), 226. https://doi.org/10.1364/JOSAA.1.000226

Ramon, M., Vizioli, L., Liu-Shuang, J., & Rossion B. (2015). Neural microgenesis of personally familiar face recognition. PNAS, 112, DOI: 10.1073.

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147.

Sinha, P., & Poggio, T. (1996). I think I know that face …. Nature, 384/6608.404.

Sinha, P., & Poggio, T. (2002). United we stand: The role of head-structure in face recognition. *Perception, 31*, 133.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologia Attention Performance, 30*, 276–315.

Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology, 28*(3-4), 156–208.

Swystun, A. G., & Logan, A. J. (2019). Quantifying the effect of viewpoint changes on sensitivity to face identity. *Vision Research, 165*, 1–12.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*(12), 1761–1771.

Troje, N. F., & Bülthoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research, 38*(1), 79–89.

Valentin, D., Abdi, H., & Edelman, B. (1997). What Represents a Face? A Computational Approach for the Integration of Physiological and Psychological Data. *Perception, 26* (10), 1271–1288.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research, 19*(1), 19–32.

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research, 42*(27), 2909–2923.

Wilson, H. R., McFarlane, D. K., and Phillips, G. C. (1983) Spatial frequency tuning of orientation selective units estimated by oblique masking. Vision Research, 23, 873–882.

Wiskott, L., Fellous, J.-M., Kuiger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions Pattern Analysis and Machine Intelligence, 19*(7), 775–779.

Xu, X., & Biederman, I. (2014). Neural correlates of face detection. Cerebral Cortex, 24, 1555–1564.

Xu, X., Biederman, I., & Shah, M. S. (2014). A neurocomputational account of the face configural effect. Journal of Vision, 14, 1–9.

Yue, X., Biederman, I., Mangini, M. C., Malsburg, C. V., & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. Vision Research, 55, 41–46 (2012).

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*, 3802–3811.