Why is it so Difficult to Recognize Faces Differing Only Moderately in Orientation in Depth?

Catrina M. Hacker[1], Tianyi Zhu[2], Miles Nelken[1], Emily X. Meschke[1], Irving Biederman[1,2,*]
[1]Program in Neuroscience, [2]Department of Psychology, University of Southern California

Attempting to recognize an unfamiliar face at an orientation in depth that differs only moderately from its previously viewed orientation is surprisingly difficult. No general account of these costs have been offered. We assessed the effects of orientation disparity in a match-to-sample paradigm of a display of three faces, with one of the two test faces physically identical to the sample, the other being a foil but both differing by 0° to 20° from the sample. The similarity of the images were scaled by a model based on simple cell tuning that correlates highly with psychophysical similarity. Two factors that produced approximately additive effects on reaction times were sufficient to account for performance on this task: 1. The orientation disparity between the matching face and the sample, and 2. the similarity of the foil to the matching face. An orientation disparity between sample and matching faces of 20° yielded a massive 301 msec increase in RTs. Such a modest orientation disparity produces a previously unappreciated enormous increase in the metric dissimilarity between matching and sample faces to a level that is equivalent to the image dissimilarity between two faces at the same orientation, but differing in race, sex, and expression.

Keywords: Face recognition; Matching Depth-Rotated Faces; Gabor-jet model; Face similarity; Match-to-sample.

INTRODUCTION

In the absence of salient distinguishing local features, recognition of 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[1-5]. Somewhat surprisingly, there has been little 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 and had to select which one of two lower test faces was identical in identity to the upper face.

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.

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, allowing precise specification of the discrimination challenge.

Some studies of the costs of orientation disparity on face recognition have used multiple foils[2], 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 face matching. 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 face produced by 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. The greater the dissimilarity of the sample to matching face and the smaller the dissimilarity between matching face and foil, the greater the expected difficulty in selecting the correct test face. In the absence of a principled measure of the similarity of faces, these two quantities had never, heretofore, been evaluated, either individually or in concert.

The Gabor-jet model[6,7] 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 .90s[8]. Justification for the Gabor-jet scaling of the similarity of faces, beyond its excellent predictability of face similarity, derives from Yue et al.'s experiments[9] showing that the representation of faces in FFA, a cortical area critical for individuating faces[10,11] 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. FFA was not sensitive to this variation when it varied among non-face, "blobs" (resembling teeth), designed to reflect the same low-level stimulus variations as faces. That it is the *image* similarity rather than the extraction of the underlying 3D representation of a face that is relevant to psychophysical matching was also evidenced by Troje and Bülthoff[12] who showed that the advantage of matching bilaterally symmetrical faces was reduced when the faces were illuminated by asymmetrical lighting.

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 in the absence of 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.

METHODS AND PROCEDURE

Participants. Sixty-five subjects (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 seconds. Subjects excluded on the basis of fast reaction times all had extremely high error rates. 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.

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 (Figure 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. 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 changes to the default face were made to render the differences between the faces largely ineffable as occurs with naturally similar faces[1,13,14]. The computer generation of the faces excluded the presence of local distinguishing features, such as a beauty mark or mole, 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 strong differences in standard population-defined categories such as race, age, expression or attractiveness which could have been used to select a response. All the stimuli were in grey scale and were 256 x 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 dissimilarity value of 1.50 or greater were used. For analysis, the dissimilarity values between the matching and distractor faces were divided into four bins with an equal number of trials in each bin.

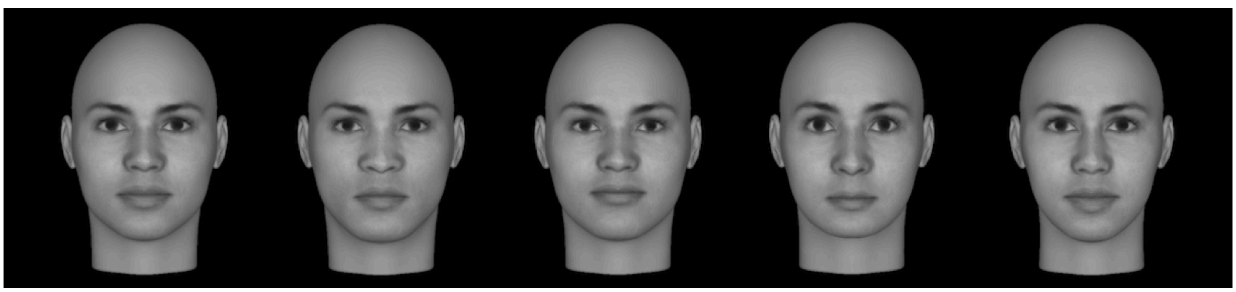

Figure 1. Figure 1. Five examples of the 20 experimental faces. Subtle variation in the features created slightly different appearing individuals. The normalized Gabor dissimilarity (Margalit et al., 2016) between faces 1 and 2 is 149 and between 1 and 5 is 363.

Figure 2. 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 between the matching and foil faces was 274 for the left panel and 293 for the right panel. The Gabor dissimilarity between the matching test face and the sample is 0 in the left panel and 557 in the right panel. In both cases the correct match to the sample is on the right and the foil is on the left.

Design and Procedure. Subjects performed the task, which consisted of 360 2AFC match-to-sample trials, on testable.org. 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 (Figure 2). 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. 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 stimuli were displayed for 5 seconds, although responses were recorded after the cessation of the display on the rare occasion that response time exceeded 5 seconds 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 minutes, which included 5 minutes for instructions.

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 (Figure 2). An important feature of the display is the diagonal arrangement of the faces which defeats a local pixel- or feature-based comparison processes which could be more readily engaged if the faces were aligned vertically or horizontally.

Stimulus similarity scaling. The Gabor Jet scaling of the physical similarity of pairs of faces was computed from a 10 x 10 grid centered on each face. (The procedure is illustrated in Margalit et al.[7] which presents an app for computing the Gabor similarity of pairs 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 x 80 kernels). Image similarity was computed as the Euclidean distance between two 8000-value vectors. The dissimilarity values reported in the present paper are normalized (by dividing by 100, the number of jets) yielding the average jet dissimilarity value for a pair of faces. When images of two faces are identical, the dissimilarity is zero.

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, with correlations in the mid .90s[8]. 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 150 to 412 which would, for most subjects, place them in an intermediate range of difficulty that allowed performance to reflect the experimental variations.

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)[15,16], in which the individual jets are not centered in a regular grid with somewhat arbitrary positioning of jets to face features, but are positioned over specific facial landmarks, 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 with 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 face to the sample under rotation so all analyses employed the grid model. In the Results section 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.

A particularly striking effect apparent in Fig. 3 is that the sizable difference in dissimilarity of about 3 Gabor units between matching 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 Gabor units from the sample), virtually disappears at 13° and 20°. 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.

Unlike the dissimilarity of matching to sample faces, the average dissimilarity between the 20 matching and foil test faces remained relatively constant over rotation angle between sample and test faces as the matching and foil faces were always presented at the same orientation. 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 as to 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.

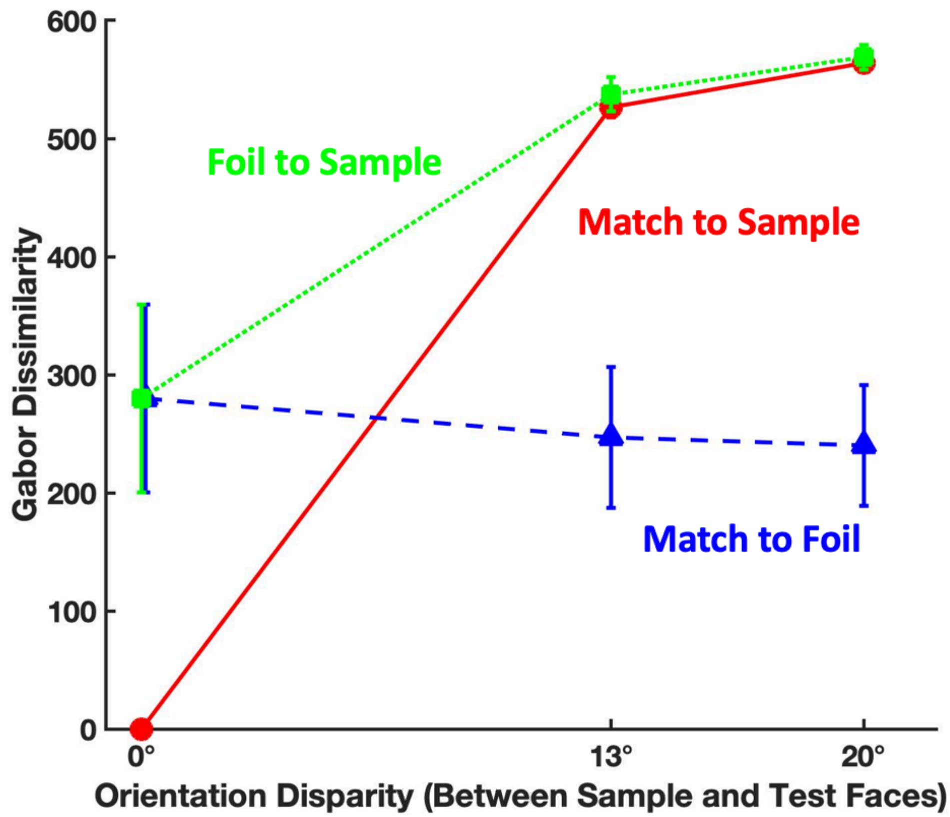

Figure 3. Mean Gabor dissimilarity values for the three stimulus relations on each trial (Match to Sample, Foil to Sample, and Match to Foil) as a function of the orientation difference between the sample and test faces. The matching and foil test faces were always presented at the same orientation. 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.

RESULTS

Effect of orientation disparity between sample and test stimuli. Figure 4 shows the mean correct RTs and error rates as a function of the angular disparity between sample and matching stimulus. For a given orientation disparity, the data are collapsed over the particular orientations of the sample and the matching test faces, e.g., 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 that variable. When matching faces at different orientations, subjects appeared more willing to tolerate longer RTs than a higher error rate. An orientation difference of 20° produced an increase in error rates of only 3.79% 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. Mean RTs were 2080, 2318 and 2381 msec for orientation differences between matching and sample faces at disparities of 0°, 13°, and 20°, $F(2,62) = 9.84$, $p < .001$, Cohen's $f = 0.31$ (a medium effect size). For error rates, $F(2,62) = 3.56$, $p < .05$, Cohen's $f = 0.19$ (a small effect).

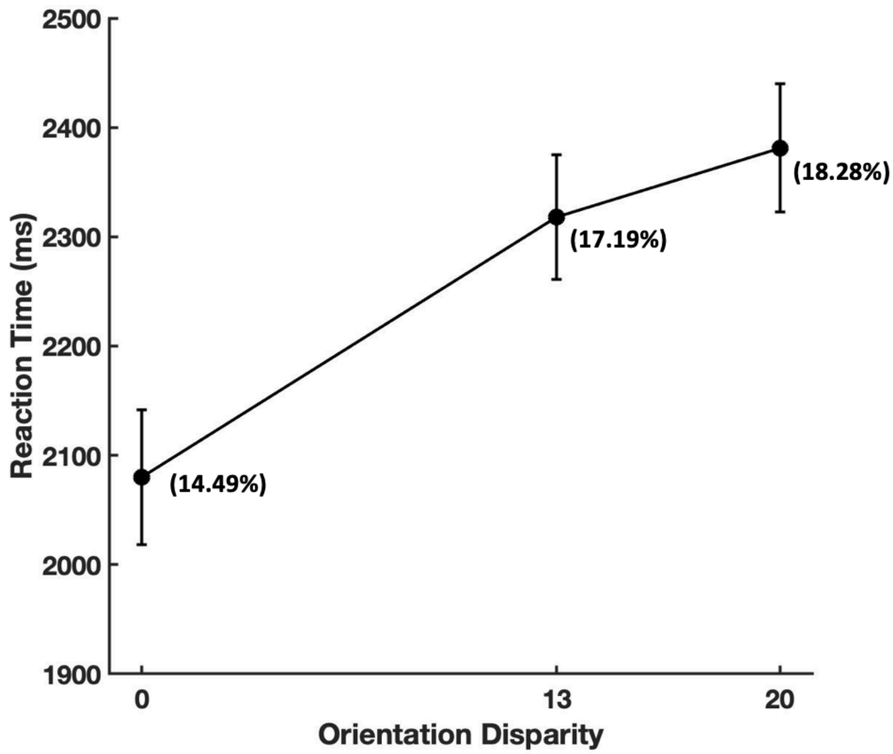

Figure 4. Mean Correct Reaction Times and Percent Errors (in parentheses) as a function of the orientation difference between matching and sample faces. Error bars are standard errors of the mean response time of the subjects.

Stimulus similarity effects. The cost of the modest 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. Over the same orientation disparities, the *mean* similarity between matching and foil faces remained largely constant 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. The results of this analysis are shown in Fig. 5. 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 and this effect is roughly independent of orientation disparity with the shallower slope of the 13° 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. 5 were plotted as inverse efficiencies in which each RT was divided by the accuracy (percent correct) at that point. The picture that emerged was vertically identical to the data in Figure 5 so differences in error rates are unlikely to be the cause of the shallower slope at 13°.

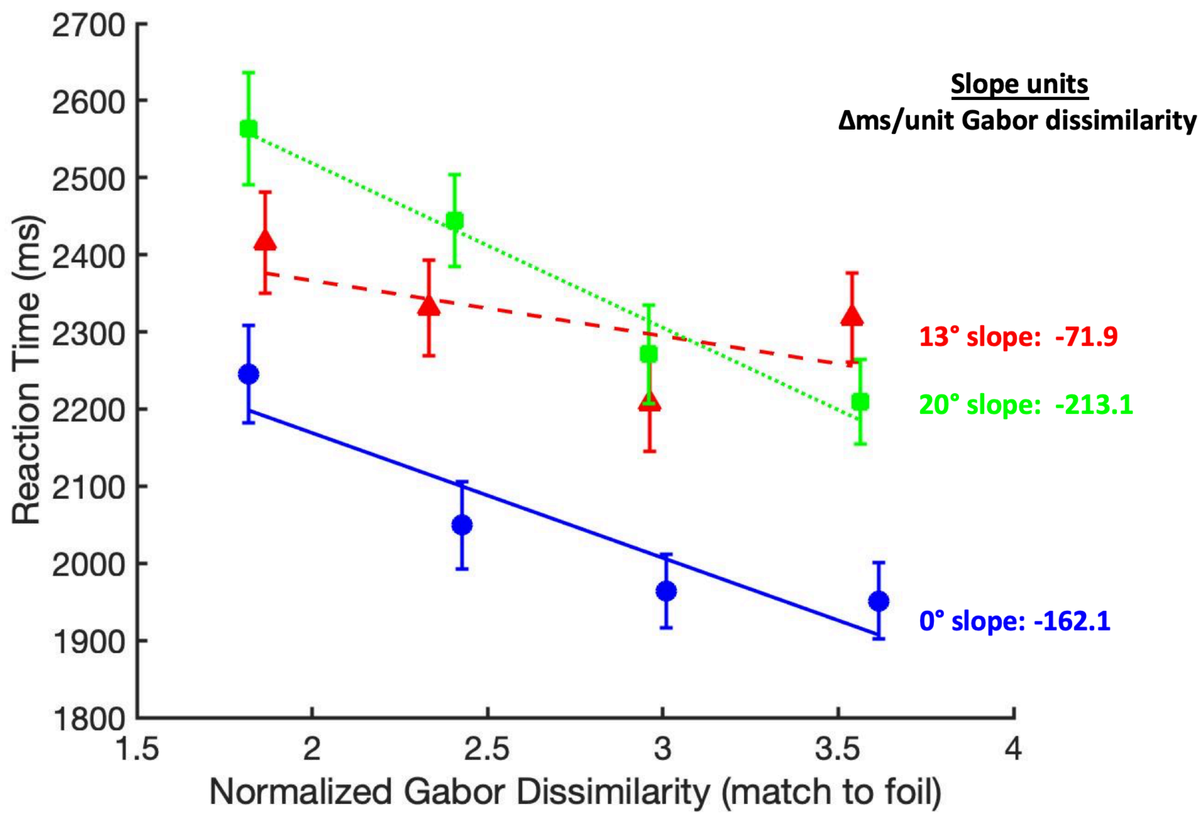

Figure 5. Mean correct reaction time (msec) as a function of the normalized Gabor dissimilarity values between the matching and the foil faces over the three levels of orientation disparity. The slopes are in units of msec per unit of normalized Gabor dissimilarity. The correlations between the Gabor dissimilarity values and RTs are -0.92, -0.61, and -0.98 for 0, 13, and 20 deg. respectively. The mean slope over the three orientations is -147.93 msec/unit of Gabor dissimilarity.

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 567 Gabor units producing a 301 msec increase in RTs or an increase of 53 msec per unit of reduced Gabor similarity. As shown in Fig. 5, the mean slope of the three functions is -1.49msec/unit of Gabor dissimilarity suggesting that the effect of similarity between the foil and matching stimulus is approximately 2.8 times the effect of dissimilarity between the matching test stimulus and the sample. Put another way, for each unit of increased Gabor similarity 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.

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° orientation 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.

It might seem surprising that the FPM in its explicit coding of aspects of face knowledge does not achieve any grain 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 world rather than in determining its identification.

DISCUSSION and CONCLUSION

Although sizable costs in recognizing or matching 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 similarity parameters are sufficient to provide an explanation 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. As scaled by the Gabor-jet model, increases in the values of each of these parameters produced linear increases in RTs. Moreover, the effects of the two were approximately additive on RTs.

The present investigation demonstrated that these modest orientation disparities produced large metric dissimilarities in the images. Figure 6 furnishes a subjective yardstick as to how a 20° orientation disparity between two computer generated faces (left panel) can be interpreted in terms of photos of familiar individuals. The dissimilarity of a 20° orientation disparity is only slightly lower than the dissimilarity of two faces at the same (0°) frontal orientation but differing in sex, race, expression, and facial hair, such as Will Smith and Angelina Jolie (Figure 7, right panel). The large costs in RTs and error rates in matching faces differing in orientation could thus be attributable to the heretofore unappreciated massive increase in the dissimilarity in images of faces produced by modest orientation disparities.

The dissimilarity of two images of the same person, i.e., the sample and the matching face, at different orientations arises from metric variations in the positions and extents of the various facial features. 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 with the difficulty in matching wire frame objects resembling bent paper clips that were studied extensively in the 1980s and 1990s[17,18]. This has also been demonstrated with faces, as in the striking demonstrations of Sinha and Poggio[19,20] in which observers fail to notice the substitution of a recent President's face for the face of his vice president. In these instances, these facial features vary metrically between the individuals and are dominated by the larger qualitative differences in hair line, head shape, glasses, and context. Similarly, the large metric differences imposed by rotation in depth that were documented in the present study produced massive costs in matching faces.

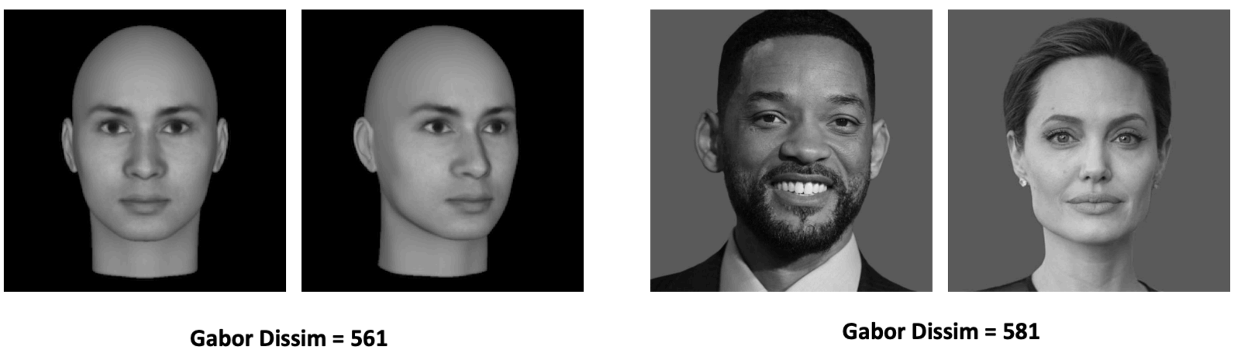

Figure 6. The dissimilarity between two of experimental faces of the same identity with an orientation disparity of as little as 20° (left panel) is about as large as two faces (right panel), at the same orientation, differing in sex, race, expression, and facial hair . Right panel, GettyImages.com.

This study documented a large cost (an increase of 301 ms in mean correct RTs) when attempting to match a face that differed from its identity match of as little as 20° from its original viewing angle. This large cost came as a surprise to most people who attempted the task. Interviews suggest that observers have knowledge about the *general* qualitative effects of rotation in depth, e.g., that one side of the face will be foreshortened and more of the other half revealed, but they appear to have difficulty in computing what the image would look like for an unfamiliar face. When we encounter new individuals, we store the images of their faces at their various orientations for subsequent recognition which is why the large costs of depth-rotated facial images are only apparent with unfamiliar faces.

REFERENCES

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

[2] 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, 576-585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

[3] Hill, H., Schyns, P., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition,62*(2), 201-222. doi:10.1016/s0010-0277(96)00785-8

[4] Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research,36*(12), 1761-1771. doi:10.1016/0042-6989(95)00230-8

[5] 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. doi:10.1068/p261271

[6] Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., Malsburg, C. V., Wurtz, R., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers,42*(3), 300-311. doi:10.1109/12.210173.

[7] 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(8), 2298-2306. doi:10.3758/s13414-016-1191-7.

[8] 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. doi:10.1016/j.visres.2011.12.012

[9] Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research,46*(22), 3802-3811. doi:10.1016/j.visres.2006.06.017

[10] Kanwisher, N. & Yovel, G. (2006). The Fusiform Face Area: A Cortical Region Specialized for the Perception of Faces. *Philosophical Transactions of the Royal Society of London B*. 361, 2109-28.

[11] 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

[12] 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. doi:10.1016/s0042-6989(97)00165x

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

[14] Xu, X., & Biederman, I. (2014). Neural correlates of face detection. *Cerebral Cortex, 24,* 1555-1564. doi: 10.1093/cercor/bht005.

[15] Wiskott, L., Fellous, J. M., Krüger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 775–779.
[16] Müeller, M. K. & Wuertz, R. P. (2009). 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.
[17] Biederman, I. (2001). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision*, 13, 241-253.
[18] Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, 4, 401-414.
[19] Sinha, P., & Poggio, T. (1996). I think I know that face …. *Nature*, 384/6608.404.
[20] Sinha, P., & Poggio, T. (2002). United we stand: The role of head-structure in face recognition. *Perception*, 31/1, 133.

The authors declare no competing interests.

We are grateful to The Dornsife Research Fund for their support.

Open Practice Statement: The data have been deposited in The Open Science Framework, osf.io.

Author Contributions

CMH: Contributed to the theoretical development of study; performed advanced data analyses and scaling of stimuli; contributed to the literature review; participated in the writing.
TZ: Supervised data collection; Organized data; Created graphs; Participated in pilot work.
MN: Created stimuli; Ran subjects; Performed early data analyses.
EXM: Participated in early phases of study; organized data; performed data analyses and fiducial point scaling of the stimuli.
IB: Conceived of and designed the study; developed the theory and guided the data analyses; primary author of manuscript.