

Recognizing Facial Slivers

Sharon Gilad-Gutnick¹, Elia Samuel Harmatz², Kleovoulos Tsourides³,
Galit Yovel⁴, and Pawan Sinha¹

Abstract

■ We report here an unexpectedly robust ability of healthy human individuals ($n = 40$) to recognize extremely distorted needle-like facial images, challenging the well-entrenched notion that veridical spatial configuration is necessary for extracting facial identity. In face identification tasks of parametrically compressed internal and external features, we found that the sum of performances on each cue falls significantly short of performance on full faces, despite the equal visual information available from both measures (with full faces essentially being a superposition of internal and external features). We hypothesize that this large deficit stems from the use of positional information about how the internal features are positioned relative to the external features. To test this, we systematically changed the relations between internal and external features and found

preferential encoding of vertical but not horizontal spatial relationships in facial representations ($n = 20$). Finally, we employ magnetoencephalography imaging ($n = 20$) to demonstrate a close mapping between the behavioral psychometric curve and the amplitude of the M250 face familiarity, but not M170 face-sensitive evoked response field component, providing evidence that the M250 can be modulated by faces that are perceptually identifiable, irrespective of extreme distortions to the face's veridical configuration. We theorize that the tolerance to compressive distortions has evolved from the need to recognize faces across varying viewpoints. Our findings help clarify the important, but poorly defined, concept of facial configuration and also enable an association between behavioral performance and previously reported neural correlates of face perception. ■

INTRODUCTION

Humans can efficiently identify hundreds of faces despite drastic changes to their appearance. Modeling this ability has attracted much research and attention. Yet fundamental questions remain unanswered, particularly regarding what spatial information is diagnostic for determining facial identity. Such questions have a long history, with rules of thumb defining the relative proportions of the head and face areas dating as far back as the Renaissance period (Cennini, 1899). These standards, like the width of the head being roughly two thirds of its height, are consistent across human faces and guide many professionals whose work focuses on faces, such as artists and plastic surgeons, as well as the development of man-machine interfaces. Indeed, configuration is believed to serve a foundational role in the domain of visual recognition, and as such, research on face perception has focused extensively on the role of spatial or configural properties, converging on the importance of the shape and/or location of internal features (McKone & Yovel, 2009; Peterson & Rhodes, 2003; Maurer, Grand, & Mondloch, 2002) and the relative contributions of internal versus external features (Longmore, Liu, & Young, 2015; Sinha & Poggio, 1996, 2002). However, if spatial

configuration is so important, and in particular consistencies in proportions, then we would expect that any distortions causing changes to the overall relative proportions would substantially disrupt recognition.

Despite the continued referencing in recent literature to the role of second-order relations or distances between internal features as playing a primary role in face processing (Tanaka & Gordon, 2011), the past few years have seen an increasing number of studies suggesting that this well-entrenched description of facial configuration cannot account for the complex performance of the human visual system (Burton, Schweinberger, Jenkins, & Kaufmann, 2015). For example, exaggerating distances between features does not impair significantly the recognition of upright faces (Caharel, Fiori, Bernard, Lalonde, & Rebai, 2006). In addition, interfeature distances seem to be less useful in similarity judgment tasks compared with local feature shape (Rhodes, 1988). Finally, the fact that both the shape and distance between facial features become variable across many normal viewing conditions, such as those resulting from facial movements during expression changes, suggest that the visual system is unlikely to rely on these spatial measurements for performing a face identification task.

Our first goal was thus to find out how much distortion in the appearance of a face the visual system can tolerate. A previous study has shown that recognition performance is tolerant to modest stretching of the face by

¹Massachusetts Institute of Technology, ²Stanford School of Medicine, ³California Institute of Technology, ⁴Tel Aviv University

up to 200% (Hole, George, Eaves, & Rasek, 2002), providing strong evidence that face recognition does not rely on the distance between individual features. However, it is unknown how performance is affected by more extreme distortions. Resolving the limit or recognition threshold of compressed faces is the first step in determining what critical facial information is lost when the system can no longer derive the identity of the face. To this end, we parametrically applied a set of nonuniform distortions to frontal facial images of famous individuals along either the horizontal or vertical axes (“thinning” or “flattening” images, respectively) and amplified these compressions to the point where the face loses any identity information and even stops resembling a face altogether. Thus, at the extremes of the compression range used, the images were reduced to needle-like facial slivers.

In addition to yielding clues about the nature of facial information used for identity judgments, compressive transformation can also serve as a tool to relate behavioral performance to known neural signals. Demonstrating that known neural markers of face processing are modulated similarly to the behavioral performance curve along the axis of compression would reinforce our behavioral findings and allow us to further break down these facial distortions to determine which facial attributes drive such neural signatures. EEG and magnetoencephalography (MEG) have led to the discovery of a discrete occipitotemporal ERP (evoked response potential) component: The N170 component that shows up in EEG measurement and its MEG counterpart, the M170 (Gao et al., 2013; Liu, Harris, & Kanwisher, 2002), occur ~170 msec after stimulus onset and are about twice as large in amplitude when the participant views a face stimulus rather than a nonface object (Liu, Higuchi, Marantz, & Kanwisher, 2000; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1996). Notably, this component seems to be related to first-order configuration (general arrangement of eyes above nose above mouth), because inverted and scrambled faces result in a delayed 170 response and, sometimes, increased amplitude (Eimer, 2000; Rossion et al., 2000; Bentin et al., 1996). Finally, many studies have demonstrated that the N170 is not affected by the familiarity of faces (Bentin & Deouell, 2000; Eimer, 2000; but see Caharel et al., 2002, for diverging results), indicating that this component is associated with face-processing stages that precede the identification of individual faces. The N170 is therefore generally thought to reflect the perceptual structural encoding of a face (Young & Bruce, 2011; Liu et al., 2002; Eimer, 2000). Shortly after this category-based component, face familiarity is believed to occur approximately 250 msec after the stimulus onset, as is reflected by the negative N250 component (Tanaka, Curran, Porterfield, & Collins, 2006; Bentin & Deouell, 2000; Schweinberger, Pfütze, & Sommer, 1995). Based on these findings, we located the M170 and M250 components while our participants

viewed both faces and objects that had been subjected to a range of compressions. Our goal was to measure how whole-image distortions modulate neural responses to known face images, but not object images, as a function of compression and to relate these curves to the behavioral performance curves. This approach provides an important link between neural activity and perception and has important implications for understanding the nature of the configural information that subserves face recognition.

Our behavioral and magnetoencephalographic studies revealed that even faces compressed so drastically as to become needle-like slivers are still easily identifiable and similarly modulate some electrophysiological face markers across participants. Building on this unexpected result, we next introduced variations of this distortion to isolate what diagnostic information may drive recognition and face-related neural markers at extreme compressions. Much of the past work on face processing has treated the mutual spatial configuration of “internal” facial features (the eyes, nose, and mouth) as being the primary determiner of identity. This assumption predicts that performance obtained for compressed “internal correct configuration” features should, on its own, largely account for the performance observed with compressed “full faces.” Thus, we determined the relative contributions of internal and external features for preserving the identity in highly compressed faces and further probed how these cues may interact to drive identification of facial slivers.

Finally, in our discussion, we hypothesize about the possible evolutionary significance of recognizing severely compressed faces and argue that tolerance to compressions is a byproduct of the brain’s need and ability to recognize depth-rotated faces. Overall, our findings question dominant accounts of face recognition that rely on veridical configural cues and lead to novel predictions about what type of information is important for preserving a face’s identity across drastic changes in appearance, such as those resulting from viewpoint changes.

METHODS

Behavioral Experiments

Full Face and the Roles of Internal and External Features

Procedure for generating stimulus set. Stimuli were generated from an original image set of 40 high-resolution, frontal, full-color images of male and female white faces of famous individuals that our lab compiled using Google images and used in previous experiments (Ehrenberg, Tsourides, Nejati, Man, & Sinha, 2017). Images excluded nonfacial cues, such as facial hair, glasses, or jewelry, and were tightly cropped to include the outline of the face and hair and then presented on a uniform white background. Image preprocessing included luminance

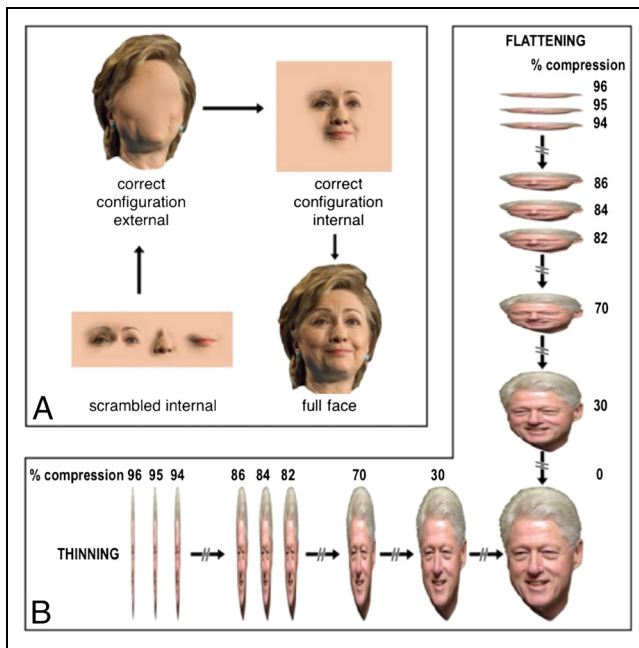


Figure 1. Four stimulus classes and sample compressions of full-face images. (A) Participants performed a naming task on compressed images from each of the four stimulus classes (here we show only uncompressed version). Each subject participated in all four conditions but saw mutually exclusive identities for each condition in a blocked paradigm, with order indicated by the arrows. (B) Trial of thinned and flattened faces. Participants saw a face that was compressed either vertically (flattening) or horizontally (thinning), beginning from the highest compression and gradually progressing to the noncompressed face. The highest compression at which each face was recognized was recorded for each participant. The same format of gradual uncompressing was used for all four conditions shown in A.

normalization, scale normalization (50-pixel interpupillary distance), and aligning of all face images at the nose tip. From each image, the following four stimulus classes were generated: scrambled internal features (placed in a row, thus devoid of mutual configuration), internal features (in their veridical mutual configuration), external features (in their veridical mutual configuration), and internal and external features together (i.e., full faces; e.g., Figure 1A for four image classes of Hillary Clinton). Next, for each of these four stimulus classes, the following 23 compression levels were generated of each individual face image, starting at the highest compression (96%) and progressing to no compression (0%): 96% → 95% → 94% → 93% → 92% → 91% → 90% → 88% → 86% → 84% → 82% → 80% → 75% → 70% → 65% → 60% → 55% → 50% → 40% → 30% → 20% → 10% → 0%. The 23 compressed images were generated once along the vertical axis (flattening) and once along the horizontal axis (thinning). Figure 1B shows a subset of compressions generated for the full-face image of Bill Clinton along both the vertical and horizontal axes. Note that the same compression levels were generated for all stimulus classes, not just full faces. Thus, for each of the 40 individual faces, a total of 184 images were generated: (4 stimulus classes) ×

(23 compression levels) × (2 compression directions). All image manipulations were performed using Adobe Photoshop CS2.

Experimental paradigm and analysis. Forty participants (18 men, 22 women, ages 18–25 years old) with normal or corrected-to-normal vision participated in the four conditions shown in Figure 1A, in accordance with the MIT ethics committee. Formal written consent was obtained for all participants. As Figure 1A shows, the four stimulus classes were presented as blocks in the following order: scrambled internal → correct configuration external → correct configuration internal → full face. Note that, because of the inherent variability between shapes of images from the different stimulus classes as well as the different face images (due to different haircuts, etc.), visual angles were not strictly consistent. However, each image spanned the full size of the presentation monitor (15-in. laptop), and participants sat at a comfortable distance from the monitor (approximately 40 cm). Note that keeping strict viewing distance and visual angle was not critical, as we previously ran a version of this experiment with different-sized images and found that image size did not significantly contribute to performance level (unpublished). Each participant saw 10 nonoverlapping celebrities from each of the four stimulus classes (40 total per participant), consisting of an equal number of vertical and horizontal compressions, such that all identities that a particular participant saw in the thinning or flattening compressions or across conditions were mutually exclusive. An image sequence consisted of seeing a face from its highest compression, slowly progressing to the less compressed version. In each trial, participants saw a blank screen followed by a 200-msec image presentation and then a blank answer screen. Participants were asked to identify the faces via naming or a description (e.g., the name of the character an actor plays), and the threshold of correct recognition was recorded for each identity (highest compression at which the participant could first recognize the face). To ensure that participants were indeed familiar with the celebrities presented in the experiment, any trial in which the participant was unable to recognize the uncompressed full-face image was excluded from further analysis. Note that this was rarely the case and only 1 of the 40 celebrity images were not recognized by multiple participants. This image was thus removed from the stimulus set used in the MEG experiment.

Union of performances obtained in the internal and external conditions at each compression level was calculated as: $(p(\text{internal}) + p(\text{external}) - p(\text{internal}) \times p(\text{external}))$. Note that we used the union rather than simple summation because summing the performance levels could have led to erroneous double counting (caused by individuals in the stimulus set who are recognized in both internal and external conditions).

Internal Features Compressed within Intact External Features

Procedure for generating stimulus set. Stimuli were generated using the same original stimulus set of 40 famous faces and consisted of the same image preprocessing steps compression levels described above. In this case, the test condition (compressed images) consisted of a single stimulus class in which only the internal features were compressed along the 23 compression levels indicated in the previous experiment and placed within the noncompressed external features. Note that the identities of the internal and external features were never mixed. For each image, this was done once along the vertical axis and once along the horizontal axis, such that for each of the 40 individual faces, there were a total of 46 test image: (1 stimulus classes) \times (23 compression levels) \times (2 compression directions). In addition to the above-described test condition, for each of the 40 individual faces, two control conditions were generated: Full Faces NC (noncompressed full faces reference block) and External NC (noncompressed external features only). Figure 2 (top) shows sample images.

Experimental paradigm. Twenty new participants (mean age = 20 years) with normal or corrected-to-normal vision participated in this experiment in accordance with the MIT ethics committee. Formal written consent was obtained for all participants. As Figure 2 shows, the experiment consisted of three blocks, presented in the following order: External NC \rightarrow Internal Compressions Only (test block) \rightarrow Full Faces NC (refer-

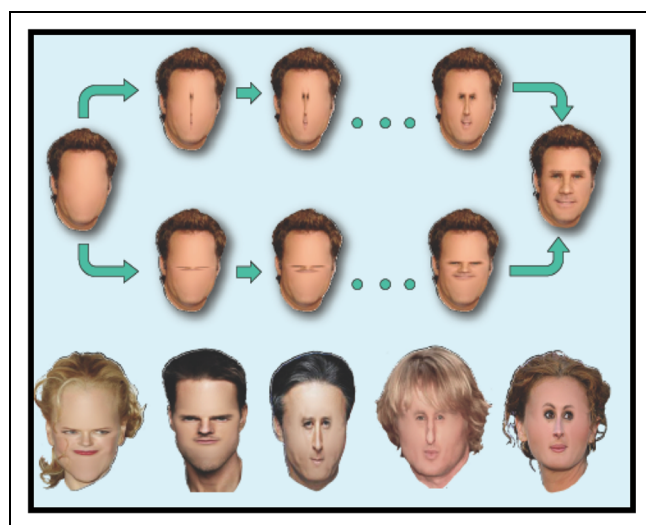


Figure 2. Sample stimuli and paradigm for internal features compressed within intact external features. Internal features were compressed either horizontally or vertically within intact external features. Participants saw three blocks: noncompressed external features only (left), internal features compressed (either vertically or horizontally) within intact external features (progressing from most compressed to noncompressed version), and noncompressed full faces (right).

ence block). Each participant saw all 40 faces on each of the above blocks and performed the same identification task as a function of compression described above on either horizontally or vertically compressed test images. To ensure that participants were familiar with each celebrity, on the one hand, but were not able to recognize the faces based on information from external features only, the data reported here includes only threshold of recognition for those test identities that were familiar (recognized in the full face third reference block) and not recognized based only on their external features (first block). Following these criteria, 2 of the 40 faces were removed completely from the analysis of the vertically compressed internal features condition.

MEG Experiment: Neural Correlates of Facial Compressions

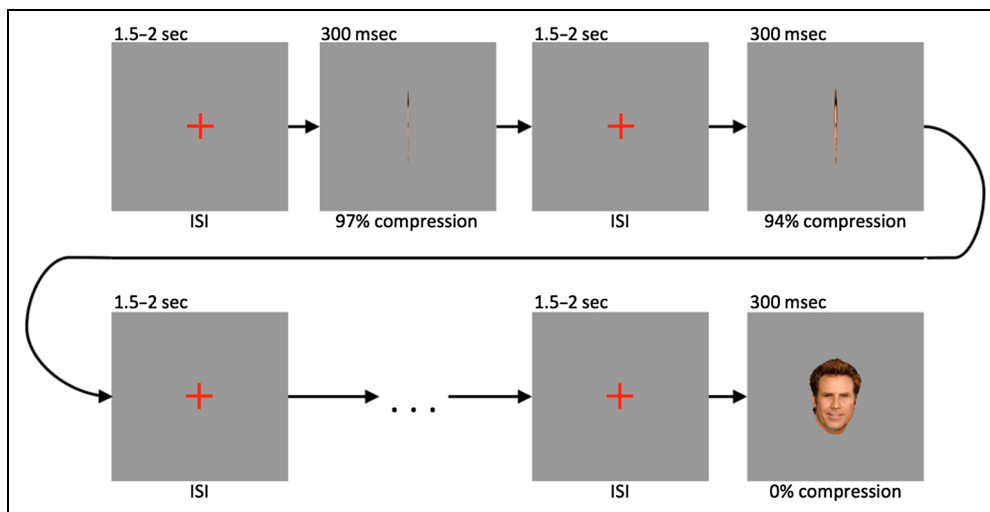
Scanner

Recordings were made using an Elekta MEG Triux scanner with 306 channels, with 204 planar gradiometer sensors and 102 magnetometer sensors. The participants comfortably sat inside the magnet while passively viewing images that were presented onto a projector screen. The lab is equipped with a high-fidelity projector (Panasonic PT-D10000U). The visual signal is projected through the wall of the shielded room to a 44-in. back-projection screen that is placed in front of the participant chair. The experimenter sat outside the shielded room and monitored the participant's attention and progress via video as well as communicated with him or her via two-way audio. Head motion was tracked using the Elekta software during signal acquisition, and eye blink removal was performed using the Elekta software from the beginning, such that only data that passed this quality check were further collected and analyzed. Preprocessing and data analysis were performed with Brainstorm software and are described further in the Analysis section.

Participants

Twelve healthy college students (mean age = 20 years) with reported normal or corrected-to-normal acuity and no history of neurological or psychiatric disorders participated in the experiment. None of the subjects participated in any of the behavioral experiments and were thus unfamiliar with the task or image set. Each participant gave written informed consent, in accordance with the MIT ethics committee. Two participants' data were excluded from further analysis: the first because no M170 was not located for this participant in the noncompressed face condition (meaning that either the participants' brain activity was not selective to faces at all or just not detectable with MEG generators for this particular participant) and the second because the participant was only familiar with 4 of the 39 famous faces presented

Figure 3. Sample image presentation sequence for MEG. All faces and objects were presented interleaved, but within a particular face or object, participants saw the face/object starting from its most compressed version, gradually progressing to the uncompressed version.



in the experiment. Thus, the data reported here are for 10 participants in total, all of whom had lived in the United States for 7 of the past 10 years (to ensure general familiarity with celebrities).

Stimulus Set

The stimulus set included 39 famous face images and 39 known objects. The face images consisted of the images used in the behavioral experiments described above (known celebrities; faces without facial hair, glasses, or jewelry). The object images consisted of household or otherwise common objects that participants would easily recognize. Both image sets were used in previous studies (Ehrenberg et al., 2017). All images were cropped from the background using Photoshop and were scale-normalized to have equal-sized bounding boxes. The two image classes were further normalized to have the same average brightness (summed brightness of all image pixels) and contrast. Finally, all images were presented on a uniform white background. Because of time constraints in using the MEG, in this experiment we generated only 10 compression levels per image that were compressed horizontally (“thinned”). Thus, for each image (both faces and objects), the following 10 compression levels were generated, starting at the highest compression (97%) and progressing to no compression (0%): 97% → 94% → 91% → 88% → 85% → 82% → 70% → 50% → 20% → 0%.

Procedure

For each participant, all faces and objects were presented in a random interleaved order. However, within a given face or object, participants saw that particular face/object from its highest compressed form and going through all 10 compression levels up to its uncompressed form. Thus, although the experiment was not blocked by iden-

tity/object, the progression of any single identity/object from most to least compressed ensured that identity-driven neural responses at a given compression level were not affected by the presentation of a previously more veridical version of the image, potentially allowing the use of top-down information. The MEG experiment did not consist of a task, other than participants being told to maintain fixation at the center of the screen. Each trial consisted of a 1.5- to 2-sec ISI central fixation cross, followed by a 300-msec centrally presented image (Figure 3). Images were back-projected on a 48 × 36 cm screen placed 140 cm in front of the participant. Within that space, horizontally compressed faces subtended on average 6° of visual angle vertically (with some variability due to hair, etc.) and 4° visual angle horizontally. Data were sampled continuously at a 1-kHz sampling rate. The location of the head was continuously monitored during the recording session by using five position indicator coils placed on the head. After the MEG experiment, each participant performed a behavioral recognition task outside the scanner to determine their familiarity with the celebrities they were presented with while in the scanner. In this task, participants were asked to name or describe each face image, presented in its uncompressed form.

Analysis

MEG data analysis was performed on signals from the magnetometers only as preliminary analysis revealed nonsignificant differences between normalized signals extracted from the gradiometers and magnetometers. Raw signals were preprocessed with the Maxfilter software (Elekta) to compensate for head movements and perform noise reduction from eyeblinks and heartbeats; raw data were preprocessed with spatiotemporal filters (Taulu & Simola, 2006; Taulu, Kajola, & Simola, 2004). This is a standard method used for artifact removal that

consists of data processing and did not result in any trial rejection. Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011) was used to bandpass filter the data at 1–200 Hz with linear phase finite impulse response to remove external and irrelevant biological noise and signal mirrored to avoid edge effects of bandpass filtering. A 60-Hz notch filter was also applied to the data and evoked response field (ERF) computations were performed in brainstorm.

For each participant, ERF components corresponding to the M170 and M250 were extracted using standard practices and included the following criteria:

- ERFs were calculated for each participant automatically using brainstorm software by finding the average peak amplitude calculated over a preselected number of sensors and across a preselected time range.
- It was ensured that the sensors selected for the M170 component were different than those found for an early M100 nonface component and identical to those found for the later M250 face familiarity component.
- To calculate ERFs, an average of five to six sensors from correct topographic locations (either left or right occipitotemporal regions) were selected for each participant individually to meet the criteria defined here.
- Left versus right hemisphere sensors were chosen on a subject-by-subject basis, based on which produced the strongest M170 for that particular participant.
- An inclusion criterion for selecting the M170 is that, in the uncompressed condition, its peak amplitude was, on average, three times as much for faces compared with objects, produced a later M250 on the same set of sensors, and was within the correct time range. Time ranges used for selecting components were 70–110 msec for M100, 120–200 msec for M170, and 200–270 msec for M250. Note that MEG can often result in earlier components, and indeed, the M170 and M250 in our experiments showed up on average ~40 msec earlier than what is normally found with EEG.

After recovering the M170 and M250 components for each participant, we recorded the peak amplitude resulting in three variables for each participant at each compression level: M170 amplitude in response to faces, M250 amplitude in response to faces, and M250 amplitude in response to objects. These values were normalized within each variable and participant, such that for each participant the vector of amplitude values for a particular component (e.g., M170 to faces) ranged from 0 to 1.

Note that the MEG procedure reported here was developed to answer the question of how previously found face-related markers are modulated by image compression level and how this relates to behavioral performance. Although we attempted to design a paradigm that would address this question, a few limitations in our design should be noted. First and most important,

study duration was limited by participant exhaustion, such that each scan session could be no longer than ~60 min. As such, we constrained the MEG experiment to thinning compressions only. The time became additionally limited because we introduced into the MEG experiment an additional stimulus class (objects) to allow extraction of face-specific M170 ERF components. As such, we needed to reduce the number of compressions per image to 10 (compared with the 23 compression levels used in the behavioral experiments). Given this constraint, we focused our sampling on the higher compression range (>80% compression) where we observed drop in behavioral recognition to potentially be able to “zoom in” to modulation of ERF responses in this range. Finally, our stimulus set is naturalistic, in that it contains both male and female images and includes hair and some variability in facial expressions and gaze. Although some may view this to be a limitation, our choice of naturalistic stimuli was motivated by our interest in processes related to real-world identification that is robust to variability between images. We believe this is a strength, rather than a weakness, of our paradigm. Indeed, performance data were highly consistent between participants and face images and across our different pilot experiments.

RESULTS AND DISCUSSION

Our first goal was to determine how much facial compression the human visual system can tolerate. Thus, in the first condition, participants viewed either a thinned or flattened “full face,” starting from the most extreme compression and progressing through 23 levels to the uncompressed face. For each participant, correct identification threshold was recorded per face (highest compression level at which participant correctly identified the face). Figure 4 shows average recognition performance as a function of compression for thinned and flattened faces (purple and pink curves, respectively). Notably, we found that face identification performance is largely invariant to compressive distortions regardless of the direction of scaling and then plummets abruptly at around 80% compression. This remarkable ability to recognize a needle-like face that is compressed to just one fifth of its original size reinforces the idea that, even with very little information, humans are able to make inferences about a face’s identity. More specifically, the compressed faces are recognized despite the fact that they do not contain information about the exact structure or location of facial features—the two most commonly proposed descriptors of facial configuration (Maurer et al., 2002; Leder & Bruce, 2000), suggesting that the spatial representations that are essential for recognition must be highly insensitive to extreme compressions. Our data also provide upper limits of compression that the recognition system can tolerate, thereby setting the stage for determining which aspects of the facial geometry are critical for identity, particularly under conditions

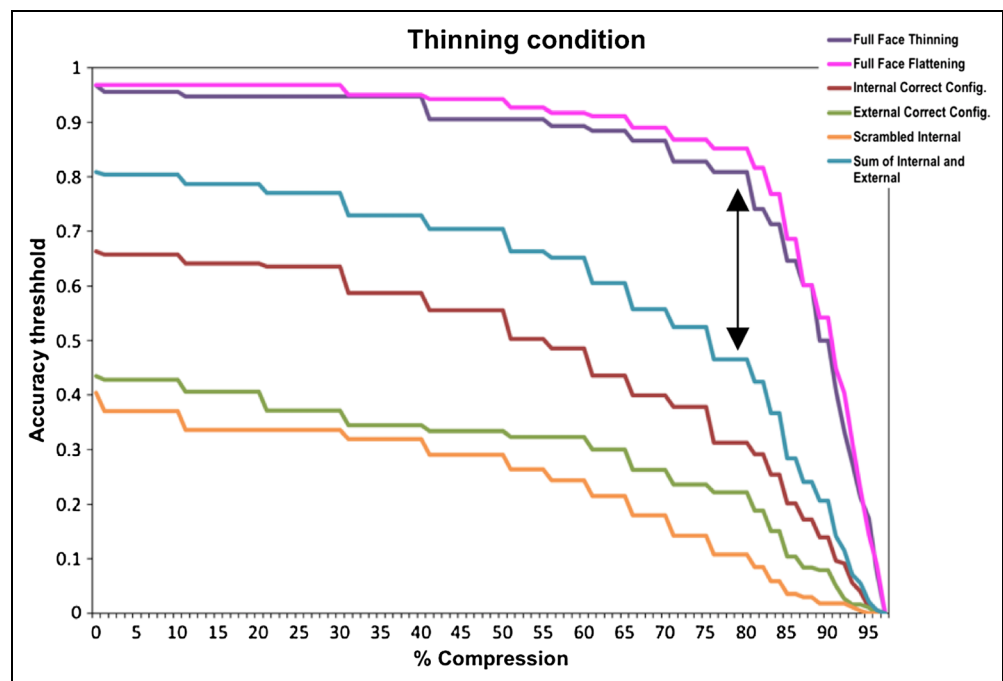
in which the appearance of a face undergoes extreme transformations, as is often the case in the real world.

Our next goal was to determine how known neural responses linked to face processing are modulated by facial compressions. We located the M170 and M250 components in the neural recordings of 10 participants, taken while they passively viewed celebrity faces and nonface objects that were subjected to a range of thinning compressions. Interestingly, notwithstanding prior reports of familiarity-driven modulations of M170 amplitude, we did not find a significant relationship between the behavioral performance curve and the amplitude of the M170 face/nonface curve (Figure 5C). However, we did observe a strong relationship between the behaviorally observed psychometric curve for identity performance and the amplitude of the M250 face familiarity marker (Figure 5A). Further validating the association between M250 amplitude and perceptual assessment of face familiarity, rather than familiarity in general, we found no systematic modulation of the M250 across the compression axis when faces were replaced by objects (Figure 5B). Overall between-participant variance on the M250 face ID marker was significantly lower than either the M170 response to faces or the M250 response to objects, both of which showed high between-subject variance (ANOVA $F(2, 18) = 20.89$, $p < .01$; paired t test: $df = 9$, $t_{(M250-Face, M250-Object)} = -5.4$, $p < .001$; $t_{(M250-Face, M170-Face)} = -5.9$, $p < .001$; $t_{(M250-Object, M170-Face)} = 1.7$, $p = .87$; Figure 5D). This high variance arises from the poor fit of the individual participants' MEG curves to the behavioral performance curve. Thus, the M250 component is induced by face images that were found to be perceptually identifiable in the behavioral experiments, despite extreme image-level distortions and

the loss of visual (and specifically configural) information in these highly compressed images. Our findings are consistent with previous studies, showing that the early N170 component is not modulated by the recognizability of a face (Caharel et al., 2006), whereas the N250r, although not sensitive to linear distortions of a face, is sensitive to the identity of the face (Bindemann, Burton, Leuthold, & Schweinberger, 2008). These studies, though informative, do not go beyond the lack of configuration to determine what type of visual information does contribute to the identity-sensitive N250 component. To do this, we performed another set of behavioral studies to determine what critical ID-specific information is lost at the threshold of recognition (around 80% compression).

To determine which specific aspects of facial geometry are perceptually significant and thus drive the above perceptual and neural responses, we applied compressive manipulations to different face parts, allowing us to tease apart what type of visual information is lost below the threshold of recognition for compressed faces. Much of the past work on face processing has treated the mutual spatial configuration of "internal" facial features (the eyes, nose, and mouth) as being the primary determiner of identity (Duchaine & Nakayama, 2006; Le Grand, Mondloch, Maurer, & Brent, 2001). This assumption predicts that performance obtained for compressed "internal correct configuration" features should, on its own, largely account for the performance observed with compressed "full faces." However, as Figure 4 (orange and green curves) shows, we found that at high levels of compression in which the full face is still easily recognizable (e.g., ~80% compression), performance on either the "internal correct configuration" or "external correct configuration"

Figure 4. Behavioral results of full-face compressions and compressions of face parts. Average accuracy thresholds for four thinned conditions and one flattened condition. Participants easily recognized highly distorted full faces (purple and pink curves), suggesting identification does not rely on veridical configural cues, and summing the performances on the internal and external conditions (cyan curve) falls significantly short of performance on the full-face condition (purple curve), suggesting critical diagnostic information is stored about where the internal features are placed relative to the external features.



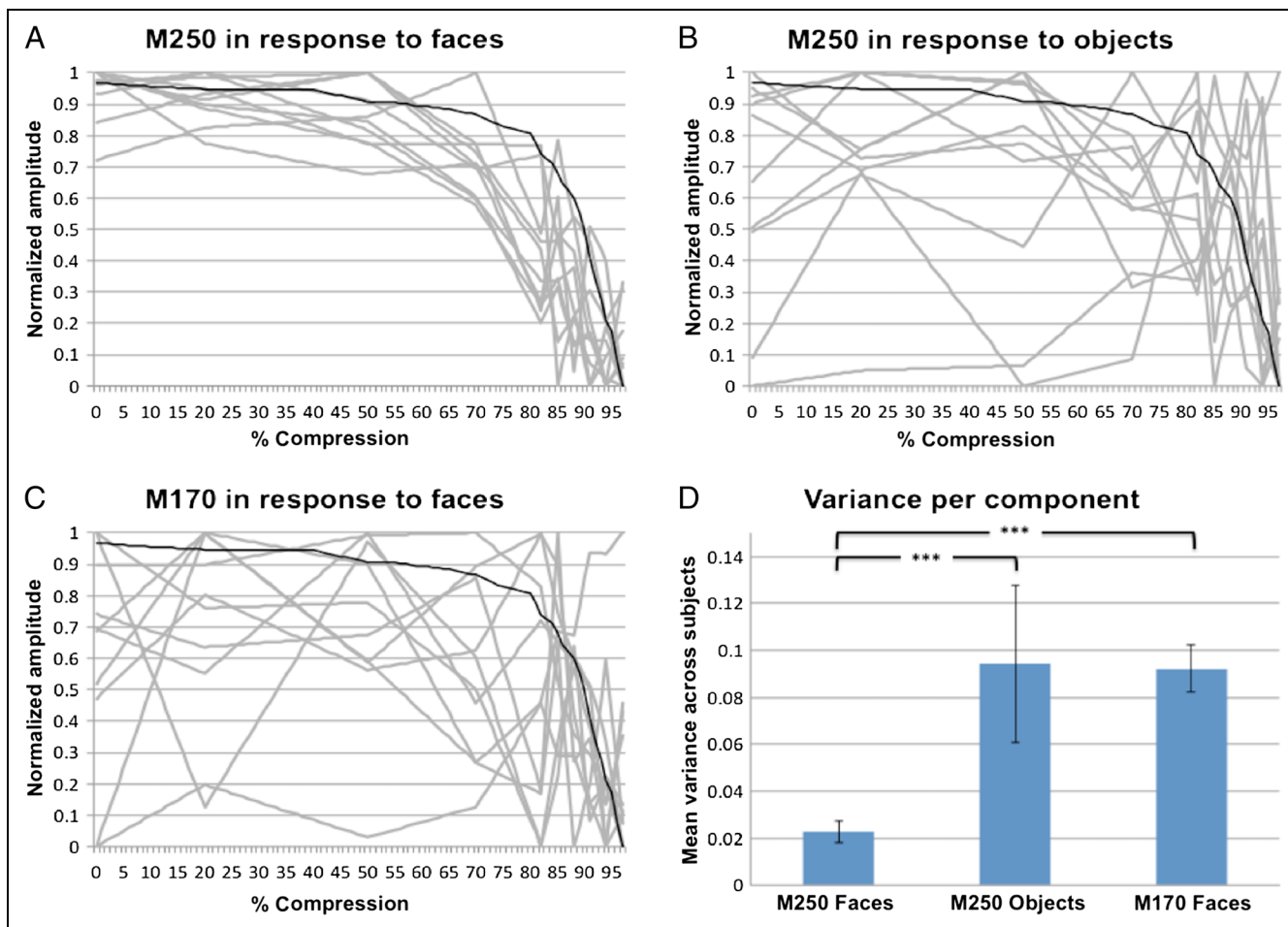


Figure 5. Individualized participants' normalized ERF responses to compressed faces and objects. Gray curves correspond to 10 participants' normalized M250 ERF amplitudes plotted as a function of increasingly compressed faces (A) and objects (B). The same 10 participants' normalized M170 ERF amplitudes are plotted as a function of increasingly compressed faces (C). In all figures (A, B, C), the black curve corresponds to the average behavioral performance curve on the full face and is shown for comparison. As the data show, only the M250 curves in response to faces are closely linked to the behavioral performance, both in size and in shape (A). This link first becomes apparent at ~80% compression or the highest compression at which the face images are behaviorally identifiable. (D) Average of between-subject variance for each neural marker. Overall between-subject variance is low only for the M250 face marker.

conditions is extremely low, suggesting that neither of these sets of features on their own can account for the robust performance obtained with the “full-face” condition.

Next, we determined how each feature type (internal vs. external features) contributes to the cumulative overall recognition of the full face. Given that, at the image level, a full-face image is essentially just a superposition of an internal and external features image, an additive cue combination approach would predict that the sum of the performances on the internal and external features conditions should be equal to the performance on the full-face condition. To test this, we computed the union of performances obtained in the internal and external conditions at each compression level (see Methods; Figure 4, cyan curve). As the difference between the purple and cyan curves shows, we found that the computed union of the performances on these two conditions falls sig-

nificantly short of the empirically observed performance on the “full-face condition” (two sample Kolmogorov–Smirnov test: $p < .01$). In fact, this deficit is most pronounced at high levels of compression, such that even when the internal and external features on their own are not informative regarding identity, presenting the two together (i.e., the full-face condition) yields a high level of performance. These behavioral data connect with the neural recordings in an interesting way. The similarity in shape between the M250 amplitude found in the MEG experiment (Figure 5A) and the behaviorally observed accuracy curve becomes apparent and consistent for all participants at 80% compression, a point at which overall configuration is still highly distorted but where we found the largest superadditive effect of presenting the internal and external features together. This suggests that the M250 may not code for overall facial configuration, as has been suggested, but rather is driven by sensitivity

to cues arising from the interaction between internal and external features. In fact, the behavioral superadditive interaction found between the internal and external features indicates that it is not sufficient for the facial representation to encode information about each feature type. Rather, it must be able to access additional information that is available only when the two feature types are presented together (e.g., how the internal features are placed relative to the external features). This idea is consistent with studies showing that the presentation of internal features can modify the neural response to external features (Axelrod & Yovel, 2010) and extends these findings by showing an additional critical role for the interaction between these two feature sets for contributing to overall face identification, particularly in highly distorted faces (the ecological significance of which we discuss in Figure 7A).

One way in which the relative placements of internal to external features can be encoded is through the use of distance ratios between the two feature types. Our results from the “full-face condition” suggest that intact “across-axis” distance ratios (such as the nose-width to head-height ratio) are not critical for preserving facial identity, because these measurements are highly distorted in our easily recognizable compressed faces. In contrast, “within-axis” distance ratios (e.g., the distance between the eyes relative to the width of the face) are preserved in the “full-face condition,” suggesting that these cues may help signal identity of the compressed faces and play a critical role for face identification in general. To test the perceptual role of within-axis distance ratios, we parametrically compressed the internal features only, either horizontally or vertically, and placed them within noncompressed external features (Figure 6A). Note that by compressing the internal features only and

leaving the external features intact, we were able generate all possible combination of within-axis distance ratios, whereas compressing both the internal and external features simultaneously would have resulted in redundant combinations of these ratios. As Figure 6B shows, there is significant perceptual importance to preserving vertical within-axis distance ratios but not horizontal ones. Even slight distortion to the distance ratios within the y -axis, but not x -axis, severely disrupts recognition performance (t test: $p < .001$). This finding significantly strengthens a line of research that emphasizes the important role of vertically arranged horizontal structures for identity mechanisms (Pachai, Sekuler, & Bennett, 2013; Dakin & Watt, 2009), because a 1-D projection of a face should preserve its horizontal structure when the internal features are horizontally compressed, but not when they are vertically compressed.

Compressed Faces as an Epiphenomenon of Tolerance to Depth Rotations

Our results reveal a remarkable resilience of the human face recognition system to extreme compressive distortions. A natural question this brings up is why we have such tolerance. Can this be explained as a learned ability, acquired through experience with printed images? Although this possibility is hard to definitively rule out, recent reports of infant and nonhuman primates’ ability to recognize spatially distorted images reduce its likelihood (Yamashita, Kanazawa, & Yamaguchi, 2014; Taubert & Parr, 2010). The asymmetry between the effects of horizontal and vertical within-axis distance ratios also challenges this idea. The alternative to individual learning is evolutionary endowment, but the need to recognize 2-D

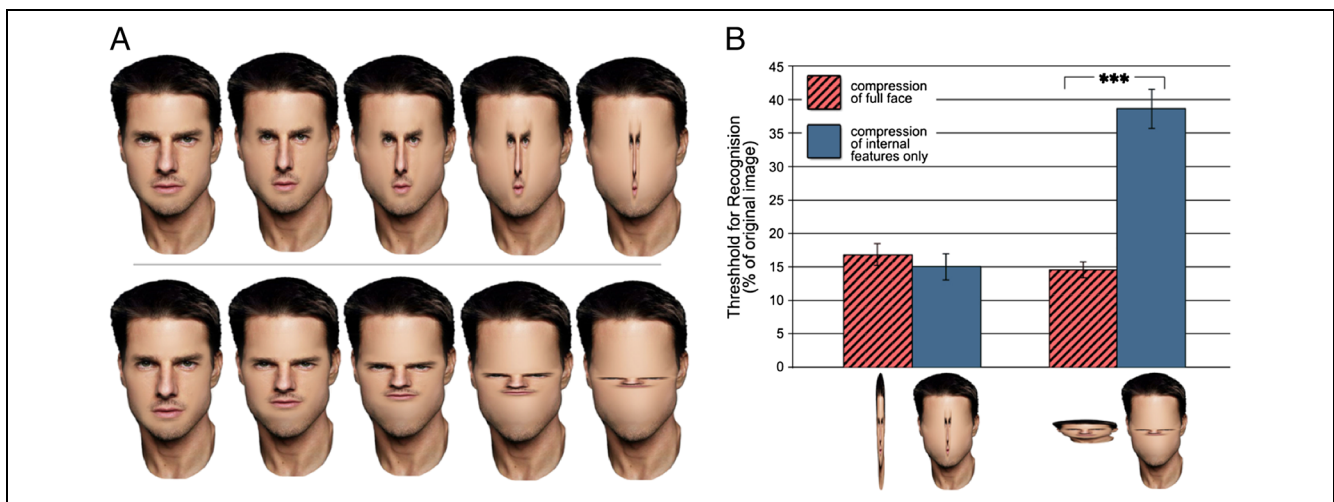
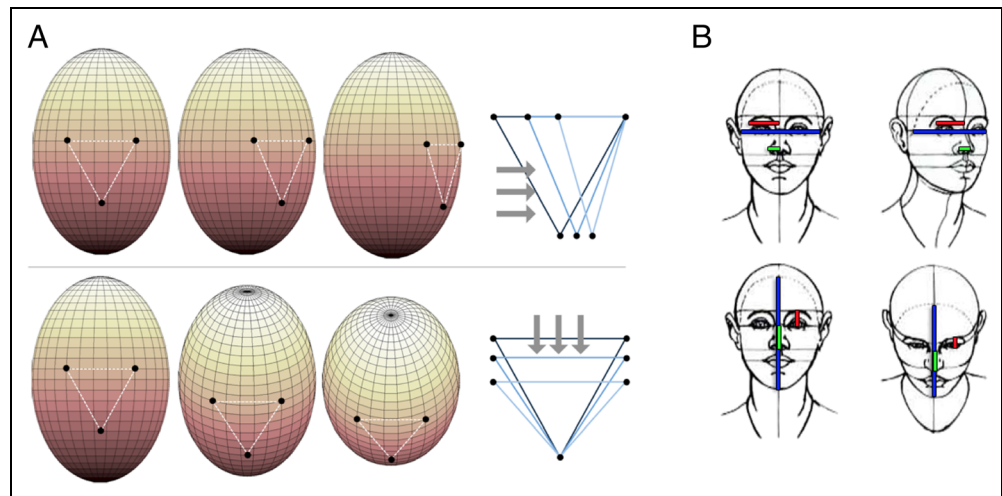


Figure 6. Sample stimuli and recognition thresholds when manipulating within-axis distance ratios. (A) Stimuli examples. Internal features were compressed either horizontally or vertically within intact external features. (B) Average recognition threshold for Block 2 reported only for faces that participant was familiar with (recognized in Block 3) but unable to recognize based on external features only (Block 1). Correct identification relies on intact within-axis distance ratios within the y axis but are not affected by severe distortions to within-axis distance ratios along the x axis.

Figure 7. The potential linkage between facial image compressions and 3-D depth rotations. (A) Geometrical transformations associated with rotating an ellipsoid (i.e., head) in 3-D space can be approximated as 2-D facial image compressions. (B) A demonstration of which within-axis distance ratios are preserved through rotation and which are not. In the frontal face, we marked distance ratios between internal and external features along the x axis (top) and y axis (bottom) and then dragged each pair to its correct location in the rotated face and preserved the original ratios between the original lines.



Distance ratios were preserved along the y axis but not along the x axis (as noted by the fact that, in the frontal image, the lines fall perfectly onto the features, whereas in the rotated image some of the lines fall short of the original feature regions). A possible explanation is that during rotations about the vertical axis the head shape (external features) remains relatively preserved, but displacement of the axis of symmetry compresses the internal feature geometry.

images of faces has not existed long enough to alter recognition mechanisms.

A more long-standing source of evolutionary pressure is the requirement to recognize faces across different viewpoints corresponding to rotations in depth. To examine how facial image information changes across depth rotations, consider the simplified depiction in Figure 7A. We assume that the head is an ellipsoid undergoing a rotation about its central axis and represents the overall layout of the face's internal features as an inverted triangle (Marquez, Ramirez, Boyer, & Delmas, 2008; Duda, Avendano, & Algazi, 1999). As the figure shows, moderate depth rotations result in the geometry of the face undergoing a 2-D compression perpendicular to the axis of rotation. Thus, to a first approximation, 3-D head rotations compress the 2-D facial image, leading us to hypothesize that tolerance to compressive distortions may have evolved from the need to recognize faces across varying viewpoints. This idea is consistent with previous claims that an algorithm built to identify faces must incorporate warping and not just aligning of face images to compensate for the changes in the distances between facial features and gain pose invariance (Martinez, 2002). To this point, Goffaux and Dakin (2010) found that horizontal structures critical for identity in general play an important role for face identification across viewpoint (Goffaux & Dakin, 2010), strengthening our hypothesis that the system's reliance on vertical within-axis distance relations may have evolved to tolerate changes in viewpoint.

The potential linkage to depth rotations also helps explain the differential significance of horizontal and vertical distance relationships for face identification. In the real-world setting, one rarely has to recognize peers across rotations about the x axis, whereas we are con-

stantly faced with changes in viewpoint about the y axis and are therefore better at recognizing faces that are rotated about the vertical axis (Favelle, Palmisano, & Maloney, 2007; Wallraven, Schwaninger, Schuhmacher, & Bühlhoff, 2002). The visual system's representation, thus, seems to only include measurements about those distance ratios that remain stable across rotation while ignoring measurements that change with viewpoint. In fact, the visual system's face representation might not encode information about distance ratios within the x axis because such cues may not be available from an actual viewed face. That is, the inherent interaction between the shape of the head and facial symmetry is such that, when the head is rotated, at least some within-axis distance ratios between internal and external features are not preserved within the x axis (Figure 7B, top) but are preserved within the y axis (Figure 7B, bottom). Thus, as viewpoint changes, the visual system would have access to intact within-axis distance ratios only along the y axis, an idea that is reinforced by our findings.

The representational biases that our data have uncovered appear to apply not just to the task of face identification but more broadly to other aspects of face processing. An unpublished study we conducted examined the effect of compressions on an ecologically important face perception task of emotion recognition. We used faces that were either flattened or thinned to 30% of their original dimension. Participants were extremely good at classifying facial expressions. These results suggest that other aspects of face processing are also tolerant to compression. The bias toward iso-dimensional relationships and away from cross-dimensional ones appears to apply to face processing broadly. Whether this bias would translate to the recognition of classes of expertise more generally (e.g., recognition of cars by car experts) is

still an open question. It is very possible that, even if faces and other classes of expertise are recognized similarly under certain viewing conditions, the strategies may diverge on certain dimensions, such as invariance to rotation.

An interesting question for future investigation is whether the response properties of neurons in inferotemporal cortex that have been reported to have rotation-invariant face responses will show invariance to the 2-D compression manipulation we have employed here. Specifically, Freiwald and Tsao (2010) found that distinct face patches within the macaque face-processing network differed qualitatively in how they responded to identity across head orientation (Freiwald & Tsao, 2010). Neurons located in the middle lateral and middle fundus regions were found to be view specific, whereas the most anterior face patch achieved almost full view invariance. Similarly, a recent fMRI decoding study found a similarly organized view-invariant face identity representation pathway in the human visual system that begins in early visual cortex and the occipital face area (OFA) with a representation of head view that is invariant to identity; proceeds to an intermediate level of representation in the face fusiform area (FFA), which represents identity entangled with head view; and culminates in the right inferior frontal cortex face area with a 3-D view-invariant representation of identity (Guntupalli, Wheeler, & Gobbini, 2017). Although currently no data exist to classify which of these areas support configural face processing, a recently published study found that the OFA and FFA are invariant to slight linear distortions, but not to nonlinear distortions (Baseler, Young, Jenkins, Mike Burton, & Andrews, 2016). This finding may initially seem counter to what our theory would predict. However, it is important to note that the linear transformations used in this study consisted of 50% compression—a distortion that may be too small for the visual system to register as being related to significant viewpoint changes. On the other hand, a pattern of activity wherein middle lateral and middle fundus patches in macaque and early visual cortex or OFA in humans are sensitive to extreme compressions at or near threshold (80% compression) whereas more anterior regions do not care about compression level but can distinguish identities at extreme compressions would further emphasize the importance of examining threshold of performance to compressions and may enable characterization of the specific mechanisms involved in coding for identity in view-invariant versus rotation-sensitive areas.

Acknowledgments

We thank Drs. Richard Held, Ming Meng, Sidney Diamond, Benjamin Balas, and Amos Gutnick for their helpful comments on this work. This work was supported by NEI (NIH) R01 EY020517 Scholar Award from James McDonnell Foundation to P. S. S. G., P. Y., and P. S. designed the study, S. G., K. T. and S. H. collected data, S. G. and K. T. analyzed data, and S. G., K. T., G. Y., and P. S. wrote the manuscript.

Reprint requests should be sent to Sharon Gilad-Gutnick, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Room 4089, 46-4089, 77 Massachusetts Avenue, Cambridge, MA 02139, or via e-mail: sharongu@mit.edu.

REFERENCES

- Axelrod, V., & Yovel, G. (2010). External facial features modify the representation of internal facial features in the fusiform face area. *Neuroimage*, *52*, 720–725.
- Baseler, H. A., Young, A. W., Jenkins, R., Mike Burton, A., & Andrews, T. J. (2016). Face-selective regions show invariance to linear, but not to non-linear, changes in facial images. *Neuropsychologia*, *93*, 76–84.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.
- Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*, 35–55.
- Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential correlates of face recognition: Geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology*, *45*, 535–544.
- Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*, 482–496.
- Caharel, S., Fiori, N., Bernard, C., Lalonde, R., & Rebai, M. (2006). The effects of inversion and eye displacements of familiar and unknown faces on early and late-stage ERPs. *International Journal of Psychophysiology*, *62*, 141–151.
- Caharel, S., Poiroux, S., Bernard, C., Thibaut, F., Lalonde, R., & Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. *International Journal of Neuroscience*, *112*, 1499–1512.
- Cennini, C. (1899). *The book of the art of Cennino Cennini: A contemporary practical treatise on quattrocento painting*. London: G. Allen & Unwin.
- Dakin, S. C., & Watt, R. J. (2009). Biological “bar codes” in human faces. *Journal of Vision*, *9*, 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.
- Duda, R. O., Avendano, C., & Algazi, V. R. (1999). An adaptable ellipsoidal head model for the interaural time difference. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings* (Vol. 2, pp. 965–968). Phoenix, AZ: IEEE.
- Ehrenberg, E., Tsourides, K., Nejati, H., Man, N., & Sinha, P. (2017). Neural correlates of face familiarity perception (unpublished manuscript).
- Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *NeuroReport: For Rapid Communication of Neuroscience Research*, *11*, 2319–2324.
- Favelle, S. K., Palmisano, S., & Maloney, R. T. (2007). Things are looking up: Differential decline in face recognition following pitch and yaw rotation. *Perception*, *36*, 1334–1352.
- Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*, 845–851.
- Gao, Z., Goldstein, A., Harpaz, Y., Hansel, M., Zion-Golumbic, E., & Bentin, S. (2013). A magnetoencephalographic study of face processing: M170, gamma-band oscillations and source localization. *Human Brain Mapping*, *34*, 1783–1795.

- Goffaux, V., & Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology, 1*, 143.
- Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the representation of identity from head view along the human face processing pathway. *Cerebral Cortex, 27*, 46–53.
- Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*, 1221–1240.
- Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition, 3*, 1–38.
- Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature, 410*, 890.
- Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 53*, 513–536.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*, 910–916.
- Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *NeuroReport, 11*, 337–341.
- Longmore, C. A., Liu, C. H., & Young, A. W. (2015). The importance of internal facial features in learning new faces. *The Quarterly Journal of Experimental Psychology, 68*, 249–260.
- Marquez, J., Ramirez, W., Boyer, L., & Delmas, P. (2008). Robust ellipsoidal model fitting of human heads. In *Robot vision* (pp. 381–390). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-78157-8_29.
- Martinez, A. M. (2002). Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*, 748–763.
- Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255–260.
- McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review, 16*, 778–797.
- Pachai, M. V., Sekuler, A. B., & Bennett, P. J. (2013). Sensitivity to information conveyed by horizontal contours is correlated with face identification accuracy. *Frontiers in Psychology, 4*, 74.
- Peterson, M. A., & Rhodes, G. (2003). *Perception of faces, objects, and scenes: Analytic and holistic processes*. New York: Oxford University Press.
- Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17*, 43–63.
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *NeuroReport, 11*, 69–74.
- Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 722–736.
- Sinha, P., & Poggio, T. (1996). I think I know that face... *Nature, 384*, 404.
- Sinha, P., & Poggio, T. (2002). “United” we stand. *Perception, 31*, 133.
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience, 2011*, 8:1–8:13.
- Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience, 18*, 1488–1497.
- Tanaka, J. W., & Gordon, I. (2011). *Features, configuration, and holistic face processing*. Retrieved from www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199559053.001.0001/oxfordhb-9780199559053-e-010
- Taubert, J., & Parr, L. A. (2010). Geometric distortions affect face recognition in chimpanzees (*Pan troglodytes*) and monkeys (*Macaca mulatta*). *Animal Cognition, 14*, 35–43.
- Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topography, 16*, 269–275.
- Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine & Biology, 51*, 1759.
- Wallraven, C., Schwanager, A., Schuhmacher, S., & Bülthoff, H. H. (2002). View-based recognition of faces in man and machine: Re-visiting inter-extra-ortho. In H. H. Bülthoff, C. Wallraven, S.-W. Lee, & T. A. Poggio (Eds.), *Biologically motivated computer vision* (pp. 651–660). Berlin, Heidelberg: Springer.
- Yamashita, W., Kanazawa, S., & Yamaguchi, M. K. (2014). Tolerance of geometric distortions in infant’s face recognition. *Infant Behavior & Development, 37*, 16–20.
- Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959–974.