

What Is a Face?

Critical Features for Face Detection

Yael Omer, Roni Sapir, Yarin Hatuka & Galit Yovel

School of Psychological Sciences

Tel Aviv University

Word Count: 3728

Correspondence regarding this manuscript should be addressed to

Galit Yovel

School of Psychological Sciences

Tel Aviv University, Tel Aviv, Israel

gality@post.tau.ac.il

Abstract

Faces convey very rich information that is critical for intact social interaction. To extract this information efficiently faces should be easily detected from a complex visual scene. Here we asked which features are critical for face detection. To answer this question, we presented non-face objects that generate a strong percept of a face (i.e., Pareidolia). One group of participants rated the faceness of this set of inanimate images. A second group rated the presence of a set of 12 local and global facial features. Regression analysis revealed that only the eyes or mouth significantly contributed to faceness scores. We further showed that removing eyes or mouth, but not teeth or ears, significantly reduced faceness scores. These findings show that face detection depends on specific facial features, the eyes and the mouth. This minimal information leads to over-generalization that generate false face percepts but assures that real faces are not missed.

Faces convey rich and valuable information including expression, eye gaze, gender, race and identity that is critical for intact social functioning. To enable fast and efficient extraction of this rich information, face processing engages a specialized processing mechanism (Kanwisher, 2000). An important initial step in face processing is the detection of a face in the visual environment (Hershler & Hochstein, 2005; Tsao & Livingstone, 2008). This process requires the extraction of features that differentiate a face from non-face stimuli. Developmental studies have shown that newborns look longer at face-like stimuli already during the first 24 hours after birth (Cassia, Turati, & Simion, 2004). They further show that early on, the visual system is not tuned to faces but to a general T-shape like pattern (a.k.a Top-heavy configuration) that is correlated with the overall structure of the internal facial features. These findings suggest the existence of an innate template that gives priority to face-like stimuli and biases the processing of information from the environment towards faces for further processing of more detailed information about expression, gaze or identity (Morton & Johnson, 1991).

The priority that is given to face-like stimuli by the visual system is also manifested by the generation of a face percept by a subset of inanimate stimuli, a phenomenon known as pareidolia (see Figure 1). These face-like inanimate stimuli were shown to elicit longer looking time relative to similar inanimate stimuli that do not generate a face percept in both humans and non-human primates (Taubert, Wardle, Flessert, Leopold, & Ungerleider, 2017). Furthermore, neuroimaging studies have shown that these stimuli generate a face-like neural response. For example, in a MEG study real faces generated a larger response than face-like inanimate stimuli at 130ms, but a similar response at 165ms after stimulus onset (Hadjikhan, Kveraga, Naik, & Ahlfors, 2009). In an fMRI experiment, the pattern of response of the FFA, but not the LOC or PPA, was correlated with the faceness of the face-like inanimate stimuli

(Wardle, Seymour, Taubert, 2017) indicating that the face-selective area is tuned to non-face stimuli that generate a percept of a face (Meng, Cherian, Singal, & Sinha, 2012).

Taken together, these findings imply that face-detection mechanisms are broadly tuned and over-generalize to a certain class of inanimate stimuli. Because these inanimate stimuli are highly variable, they can be used to answer the basic question of what are the critical features that are used by face-detection mechanisms to classify a stimulus as a face. This question cannot be answered with real face stimuli, as humans are very sensitive to all facial features of real faces and can easily detect them as faces even when the information that is provided is partial, distorted and impoverished (Sinha, Balas, Ostrovsky, & Russell, 2006). The inanimate stimuli perceived as faces, on the other hand, contain non-facial information that makes them look like a face. We therefore exploit this type of stimuli to answer the basic question of what information is needed to make a stimulus perceived as a face?

In the current study, we presented participants with inanimate stimuli and asked them to rate the degree to which they look like a face (i.e., *faceness* rating). A different group of subjects rated these same stimuli on a list of 12 facial features including local features such as eyes, mouth and teeth, as well as global features such as facial expression, symmetry or face proportion. By correlating the feature rating with the faceness rating we discovered which features are needed to perceive a stimulus as face and which are not critical. One previous study that also examined the correlation between ratings of faceness and a set of a few features (eyes, mouth, eyebrows, expression and typicality) revealed that eyes, mouth and expression predicted faceness rating (Ichikawa, Kanazawa, & Yamaguchi, 2011). However, the predictors and predicted measures were not independent, as the same participants rated all the measures. To avoid this possible influence of faceness and feature ratings, different groups of participants provided these measures of the faceness and feature rating scores (Experiment 1). In addition, in the current study we conducted a follow-up experiment

(Experiment 2), in which we removed critical or non-critical features based on the correlational results and examined the effect on faceness rating. We expect that removing critical features (i.e., features that were correlated with faceness scores) would reduce faceness rating, whereas removing non-critical feature (i.e., features that are not correlated with faceness scores) will not influence faceness rating.

Experiment 1

Methods

Participants

Seventy-two subjects participated in this study (40 women, 32 men), age range 21-44 (mean:30.61, stdv:6.06), 61 right handed and 11 left handed. This sample size was determined based on a previous similar study, which included 46 participants but collected all measures from the same sample (Ichikawa et al., 2011). To assure that feature rating was independent from faceness rating, here we collected these measures from separate groups of participants. Thirty-five participants completed the faceness rating experiment, and 37 participants completed the facial feature rating experiment. Seven participants were excluded from the facial feature rating experiment because they did not complete the second part of the experiment. One additional subject used the same rating score for all 12 features and was excluded. All participants signed a consent form. The experiment was approved by the Ethics committee of Tel Aviv University.

Stimuli

One hundred and sixteen colored images were used in this study, including photos of landscapes, houses, vehicles or furniture. The images were found on different internet websites (wtface.com, themarysue.com, reddit.com) and also by using the key words “things that look like faces” in google search engine. Of the 116 images, 86 images were perceived

as faces by the experimenters. In addition, 30 non-face images that matched the “face images” in depicting similar objects/landscapes were selected.

Procedure

Two rating experiments were performed on separate groups of participants. A faceness rating experiment in which subjects were asked to rate the faceness of the images. A facial feature rating experiment, in which participants were asked to rate the extent to which they see each of 12 facial features as detailed below.

Faceness Rating: A group of 35 participants completed the faceness rating experiment. Each of the 116 images was presented on a computer screen for 2 seconds in a random order. After the disappearance of the image a scale appeared and subjects were asked to rate the extent to which they detected a face in the photo, on a scale of 1 (there was no face in the photo) to 7 (there was definitely a face in the photo). The experiment lasted about 10 mins.

Facial Feature rating: 37 participants completed the facial feature rating experiment. Each of the 116 images was presented on the screen in a random order. The subjects were asked to rate on a scale the extent to which they can detect each of the following 12 facial features: eyes, mouth, nose, forehead, teeth, ears, eyebrows, hair, frame, proportion, symmetry and emotional expression. The scale ranged from 1 (the feature does not exist in the photo) to 7 (the feature definitely exists in the photo). The instructions stated that each feature can be either present or absent in any of the images, and that features that are clearly seen in the image should be scored 7, while a less evident feature should get a lower rating. The instructions also included clarifications with respect to the features frame, proportion and symmetry. Frame – If you see any facial features, do they appear inside any sort of frame? Proportion – If you see any facial features, are their size and distance from one another is similar to an ordinary face?; Symmetry – is the object that is perceived as a face in the image symmetric?

Each image remained on the screen till participants completed rating all the features. Because the feature rating experiment was long, it was divided into two parts, each part included half of the photos. The subjects were allowed to take a break between the two parts as long as they needed. Each part took about half an hour to complete.

Data analysis

The rating scores were converted to Z-scores within each subject for the faceness score and for each feature separately. The Z-scores were averaged for each of the 116 images across subjects, so each image had one averaged Z-score faceness rating and one averaged Z-score for each of the 12 features. Two subjects used the same rating score for one feature for all 116 images, one for 'frame' and one for 'ears'. Those features were excluded.

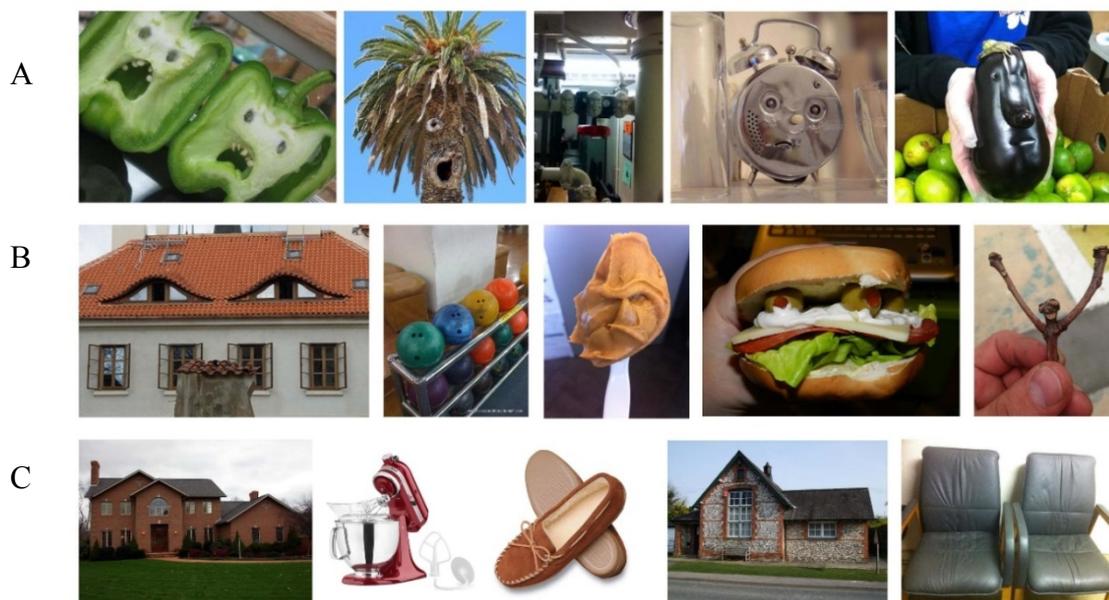


Figure 1: Inanimate images that received the 5 highest scores (A) 5 intermediate scores (B) and 5 lowest scores (C) for the question *do you see a face in the image* (i.e., faceness score).

Results

Figure 1 shows example images that received the highest, the intermediate or the lowest faceness scores. To determine which features are important for face detection we computed the correlations across the 116 images of each of the facial features rating with the faceness rating.

	Mouth	Expression	Proportion	Symmetry	Forehead	Frame	Eyebrows	Nose	Hair	Teeth	Ears	Faceness
Eyes	0.88	0.33	0.58	0.31	0.30	0.36	0.24	0.63	0.80	0.72	0.86	0.92
Mouth		0.90	0.81	0.70	0.58	0.66	0.33	0.34	0.25	0.35	0.26	0.90
Expression			0.79	0.64	0.61	0.64	0.44	0.38	0.34	0.33	0.32	0.90
Proportion				0.83	0.74	0.80	0.47	0.43	0.42	0.20	0.39	0.84
Symmetry					0.55	0.71	0.28	0.33	0.33	0.18	0.36	0.70
Forehead						0.67	0.52	0.58	0.56	0.24	0.48	0.67
Frame							0.30	0.34	0.36	0.17	0.33	0.66
Eyebrows								0.44	0.28	0.24	0.50	0.46
Nose									0.39	0.16	0.48	0.41
Hair										0.15	0.30	0.37
Teeth											0.28	0.32
Ears												0.32

Figure 2: The correlation matrix between the 12 features and the faceness score sorted based on the correlations between each feature and the faceness score (right column). The color coding indicates the strength of the correlation. Green correlations are the highest, red correlations are the lowest.

To determine the unique contribution of each of the features to faceness rating we performed multiple regression in which rating of all 12 features was used to predict rating of the faceness score. The regression model explained 92% of the variance in faceness scores. Of the 12 features, eyes and mouth contributed mostly and significantly to faceness rating ($p < .004$ corrected for 12 multiple comparisons). Eyebrows, the hair and the emotion also had a marginally significant contribution to faceness rating (see Table 1). These findings are consistent with Ichikawa and colleagues (2011) who revealed that the eyes, were the best predictor for faceness scores and mouth and expression contributed as well. In this study,

however, ratings of faceness and features were collected from the same group of participants and therefore the predictors were not measured independently.

Table 1: Results of a multiple regression of the 12 facial features and faceness rating (sorted by t-value).

	B	Std. Error	Beta	t-value	p-value
Eyes	0.50	0.07	0.49	7.05	< .001
Mouth	0.27	0.09	0.27	3.15	0.002
Eyebrows	0.13	0.05	0.10	2.47	0.015
Hair	0.10	0.04	0.08	2.21	0.029
Emotional expressiveness	0.14	0.09	0.12	1.50	0.136
Nose	0.04	0.04	0.04	0.96	0.338
Proportion	0.09	0.13	0.06	0.70	0.485
Forehead	0.03	0.06	0.03	0.52	0.605
Frame	0.00	0.07	0.00	0.03	0.978
Teeth	-0.01	0.04	-0.01	-0.14	0.887
Symmetry	-0.04	0.07	-0.03	-0.51	0.611
Ears	-0.06	0.04	-0.06	-1.47	0.145

Another measure that may be correlated with faceness scores is the number of features that are present in any of the images. An image with a higher number of features is likely to be correlated more with faceness scores than an image with a lower number of features. We therefore examined whether this variable accounts for the correlations that we found for the eyes and the mouth. For each image, we calculated the number of features that received a z-score higher than 0, indicating the presence of this feature in the image (see supplementary Tables S1, S2). The correlation of the number of features with faceness score was 0.86, $p < .0001$. We then included this variable in the regression analysis with the 12 features. This model accounts for the same proportion of variance (92%) of faceness scores as the model that did not include the number of features. Eyes (beta = .48, $t = 6.7$, $p < .001$) and mouth

(beta = .25, $t = 2.9$, $p = .005$) were still significant predictors of the faceness rating score (see full regression model in supplementary Table S3), indicating that their contribution to face detection was not mediated by the number of features present in the image.

To provide further support for the results that eyes and mouth are critical for the detection of a stimulus as a face, we created a new set of images in which we removed the critical features - either the eyes or the mouth. In addition, we created another set of images in which we removed non-critical features - either the Ears or the Teeth. Figure 3 shows example of these type of figures. These images were replaced with their original version and were presented in a faceness rating study, which included all the stimuli that were included in the previous experiment to have a similar context in both experiments. We compared the faceness scores for the stimuli with and without the removed features.

Experiment 2

Participants

99 participants volunteered to participate in the experiment for course credit. They were randomly assigned to one of the five experiments: 20 participants were presented with the ears removed stimuli (3 males, mean age = 21); 19 participants were presented with the eye-removed stimuli (3 males, mean age = 23), 20 participants were presented with the mouth-removed stimuli (2 males, mean age = 23.5); 20 participants were presented with the teeth-removed stimuli (2 males, mean age = 21.5); 20 subjects were presented with the original stimuli (6 males, mean age = 22.5), replicating Experiment 1. All participants signed a consent form. The experiment was approved by the Ethics committee of Tel Aviv University.

Stimuli

From the image set used in Experiment 1 we removed eyes from 11 images (Figure 3A), the mouth from 9 images (Figure 3B), the ears from 7 images (Figure 3C) and the teeth from 9 images (Figure 3D). We selected images that enabled the removal of a feature in a way that

does not distort or modify the original image in any way but the removal of that specific feature (see Figure 3 for examples of stimuli with and without the feature).

Methods

The experiment was similar to the previous faceness rating experiment but included a subset of images in which one of the features was removed. Participants were asked to rate the faceness of 116 images.

Data Analysis

Similar to Experiment 1, the ratings were standardized within each participant and were averaged across participants to obtain a faceness score for each image.

Results

Paired t-tests were performed to compare the faceness scores for images with and without the removed feature. Table 2 shows the original faceness scores and the scores of the same images without the removed feature. Removing the Eyes and Mouth significantly reduced the faceness scores (see Table 2). Removing teeth and mouth did not change faceness scores significantly. These findings therefore show the importance of both the eyes and mouth for detecting a face. When one of them is removed, faceness ratings are reduced significantly.

Table 2: Averaged faceness scores (standardized), effect size and paired t-tests for images with and without one feature.

	Original	Without feature	t-value	p-value	Cohen's d
Eyes	0.70	-0.21	4.94	0.002	1.28
Mouth	0.72	0.16	4.42	0.003	1.27
Teeth	0.48	0.32	1.93	0.09	0.59
Ears	0.51	0.55	0.50	0.63	-0.19

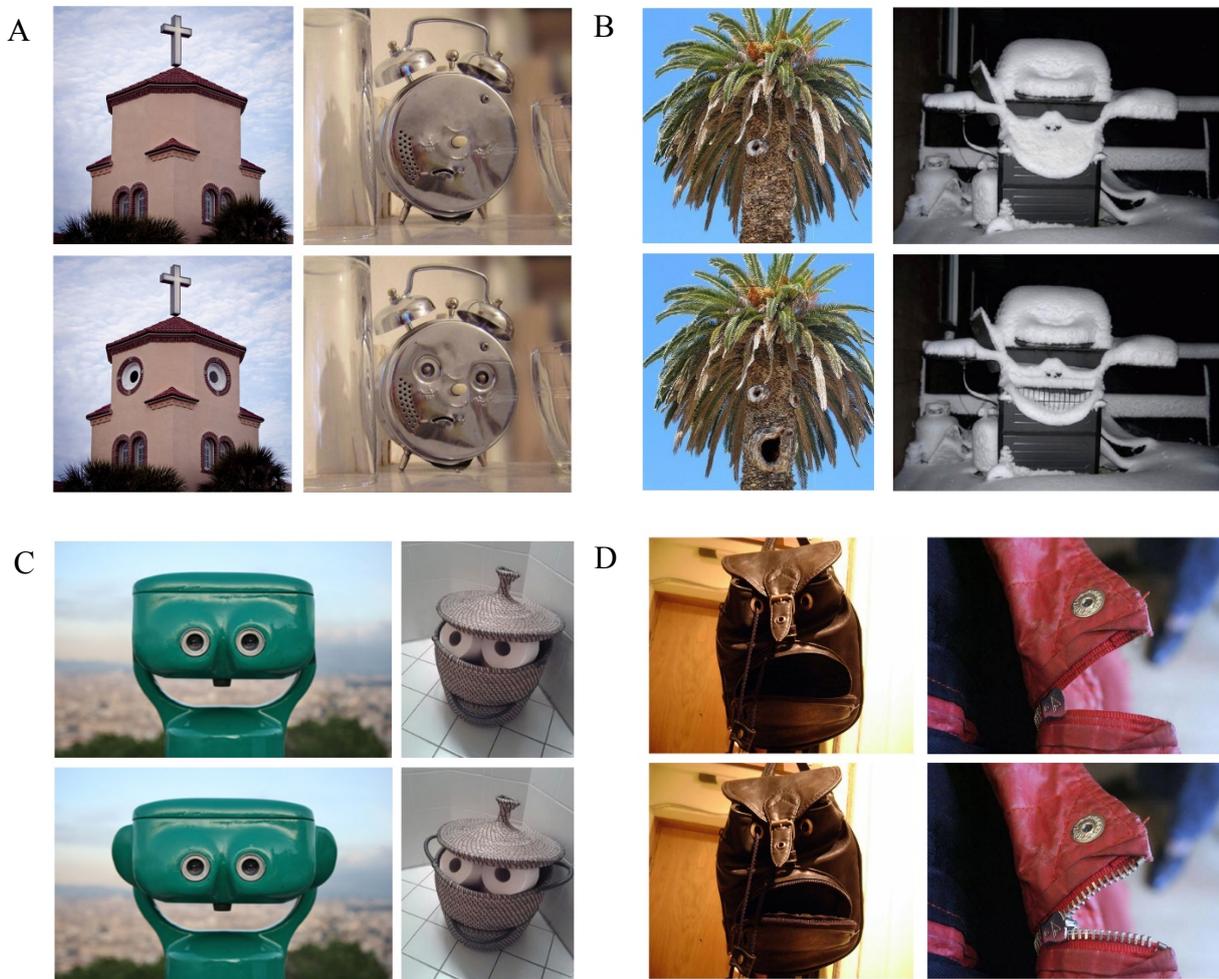


Figure 3: Stimuli used in the four versions of Experiment 2 in which a critical feature (A. Eyes; B. Mouth) or a non-critical feature (C. Ears; D. Teeth) were removed from the stimuli. Participants rated these images for faceness together with the stimuli that were presented in Experiment 1 so the same context is used for the ratings of the images with and without the feature.

General Discussion

The goal of this experiment was to discover which features are critical for face detection. To answer this question, we used inanimate images that generated a strong percept of a face and examined the correlations across stimuli between the presence of 12 facial features and faceness scores based on human ratings. Results showed that the presence of the eyes or the

mouth in the image is highly correlated with its faceness score. To further examine whether eyes and mouth are indeed critical for face detection, in a second experiment we removed either the eyes or the mouth, or two features that were not correlated with faceness, the ears or the teeth, and measured faceness scores for the edited images. Results showed that removing the eyes or the mouth significantly reduced faceness scores, whereas faceness scores of images following removal of the teeth or the ears was not significantly different from faceness scores of the original images. These findings confirm and extend the correlational findings, further showing that perceiving a face in an image depends on the presence of the eyes or the mouth in the image.

One limitation of Experiment 2 is that different images were selected for the removal of different features. This was necessary because most images did not have all the features that were selected for removal or in some cases removal of a feature distorted the image in a global way. Nevertheless, each picture was compared to itself, thus reducing the impact of this limitation. Furthermore, the faceness score of images after the removal of the non-critical features was much higher than the score for images after the removal of the critical features. Thus, removal of critical features resulted in a lower faceness score than removal of non-critical features, indicating that the smaller effect of removal of non-critical features was not due to floor effect.

Correlation analysis among the facial features (Figure 2) revealed high correlations between the eyes and mouth indicating that most images in our database that had eyes also had a mouth or had none of them. In Experiment 2, we were able to examine the contribution of the eyes and the mouth separately and found that removing one of them significantly reduced faceness scores. These findings are consistent with the regression model that estimates the unique contribution of each of the features. These findings suggest that images with either

mouth or eyes do not generate a strong percept of a face, but it is the presence of both of these features that is required to generate a face.

The importance of the eyes and the mouth for face detection in face-like inanimate stimuli was recently shown in a study with human and monkey participants that revealed increase looking time to inanimate stimuli that are perceived as faces than non-face stimuli (Taubert et al., 2017). This study also reported increased fixations to the eyes and mouth regions of these stimuli, thus indicating the central role of these features in the percept of a face. In another study, the same stimuli were presented to monkeys with amygdala lesions (Taubert et al., 2018). Results showed no bias towards faces than non-face stimuli for both real and inanimate faces. Furthermore, eye movements were not directed to the eyes and mouth for both of the face stimuli. These findings suggest an important role of the amygdala in directing attention towards faces, which is generalized also to inanimate stimuli based on the same subset of critical features.

Previous developmental studies have indicated that newborns look longer at figures with top heavy configuration such as a T-shape than an inverted T like images. Our findings that the critical features are eyes and mouth may correspond to this general T-shape pattern showing that the highest faceness scores were mostly correlated with the presence of the eyes and the mouth in the image. The current study, however, primarily focused on the presence or absence of facial features. Face detection also depends on certain spatial relations among features (a.k.a. first order configuration) (Maurer, Grand, & Mondloch, 2002) that was not manipulated in the current study. The list of features did include proportion, symmetry and frame, but these features refer to second-order relations among features. The role of face-like second order relations was examine recently by Vuong and colleagues, who interestingly found enhanced matching performance for non-face stimuli that presented face-like spatial

relations of features (Vuong et al., 2017). Thus, future studies may modify the relative locations of the critical features that we revealed to examine its influence on faceness scores. Previous studies employed other methods to determine which features are critical for face perception. For example, Schyns and colleagues (2002) developed the *bubble* technique to reveal what minimal information is needed to determine the gender, expression or identity of a face (Schyns, Bonnar, & Gosselin, 2002). Mangini and Biederman (2004) added different levels of noise to face stimuli and measured the noise patterns under which participants can determine gender, expression or face identity (Mangini & Biederman, 2004). Sekuler and colleagues (2004) also added noise to determine which minimal information is used to determine face identity for upright and inverted faces (Sekuler, Gaspar, Gold, Bennett, & Ls, 2004). Our study used a very different approach and rather than concealing different areas of the face, asked subject to evaluate the presence of its different features and examine their correlation with rating of the whole face. The disadvantage of our technique is that it requires pre-selection of features, whereas the above-mentioned methods can reveal any feature, including features that have no semantic meaning. However, our method allows us to go beyond local features and to examine the contribution of features such as proportion, symmetry or frame. Furthermore, it allows us to assess the contribution of each feature in the context of the entire face, rather than the role of local features in the absence of information from the rest of the face.

In a recent study we discovered facial features that are critical for the identity of the face (Abudarham & Yovel, 2016). These features include the hair, eyebrow thickness, eye size, eye color and lip thickness. Similar to the method used in the current study, Abudarham and Yovel (2016) showed that replacing these features with features from faces of different identities, changes the identity of a face, whereas changing other features such as mouth size or eye distance, did not change the identity of a face (see also, Abudarham, Shkiller, &

Yovel, 2019; Abudarham & Yovel, 2018). Interestingly, the top features that mostly contributed to faceness scores were the eyes, mouth, hair and eyebrows (see Table 1). Whereas features like proportion and nose, that are not critical for face identification, also did not contribute to faceness scores. This correspondence is surprising given that face detection depends on features that are shared among all face stimuli, but not with non-face stimuli, whereas face identification depends on features that are different between different identities but shared within an identity. It is noteworthy that faceness scores were correlated with the presence of these features whereas face identification depends on the specific characteristics of these features such as size, length or brightness. These findings suggest that a system that is tuned to the presence of these features as well as to detailed measurements of their shape or brightness can perform both detection and identification.

In summary, we revealed that face detection depends on a broadly tuned mechanism that classify an image as a face based on relatively minimal information that corresponds to the presence of the eyes and the mouth in the image. Such a template may generate false face percepts but assures that a real face will not be missed in the dynamic, cluttered and complex visual environment that we live in.

Acknowledgments

We would like to thank Naphtali Abudarham for his feedback on this manuscript.

References

- Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, 182(April 2018), 73–83. <http://doi.org/10.1016/j.cognition.2018.09.002>
- Abudarham, N., & Yovel, G. (2016). Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision*, 16(3), 40.

<http://doi.org/10.1167/16.3.40>

Abudarham, N., & Yovel, G. (2018). Same critical features are used for identification of familiarized and unfamiliar faces. *Vision Research*, (January), 1–7.

<http://doi.org/10.1016/j.visres.2018.01.002>

Macchi, C. V., Turati, C., & Simion, F. (2004). Can a Nonspecific Bias Toward Top-Heavy Patterns Explain Newborns Face Preference? *Psychological Science*, 15(6), 379-383.

doi:10.1111/j.0956-7976.2004.00688.x

Hadjikhani, N., Kveraga, K., Naik, P., & Ahlfors, S. P. (2009). Early (M170) activation of face-specific cortex by face-like objects. *NeuroReport*, 20(4), 403-407.

doi:10.1097/wnr.0b013e328325a8e1

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, 45(13), 1707-1724. doi:10.1016/j.visres.2004.12.021

Ichikawa, H., Kanazawa, S., & Yamaguchi, M. K. (2011). Finding a face in a face-like object. *Perception*, 40(4), 500–502. <http://doi.org/10.1068/p6926>

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3(8), 759-763. doi:10.1038/77664

Mangini, M., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209-226.

doi:10.1016/j.cogsci.2003.11.004

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255-260. doi:10.1016/s1364-

6613(02)01903-4

Meng, M., Cherian, T., Singal, G., & Sinha, P. (2012). Lateralization of face processing in the human brain, (January), 2052–2061. <http://doi.org/10.1098/rspb.2011.1784>

- Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98(2), 164-181. doi:10.1037//0033-295x.98.2.164
- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show Me the Features! Understanding Recognition From the Use of Visual Information. *Psychological Science*, 13(5), 402-409. doi:10.1111/1467-9280.00472
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion Leads to Quantitative, Not Qualitative, Changes in Face Processing. *Current Biology*, 14(5), 391-396. doi:10.1016/j.cub.2004.02.028
- Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11), 1948–1961. <http://doi.org/10.1109/JPROC.2006.884093>
- Taubert, J., Flessert, M., Wardle, S. G., Basile, B. M., Murphy, A. P., Murray, E. A., & Ungerleider, L. G. (2018). Amygdala lesions eliminate viewing preferences for faces in rhesus monkeys. *Proceedings of the National Academy of Sciences*, 115(31), 8043-8048. doi:10.1073/pnas.1807245115
- Taubert, J., Wardle, S. G., Flessert, M., Leopold, D. A., & Ungerleider, L. G. (2017). Face Pareidolia in the Rhesus Monkey. *Current Biology*, 27(16), 2505-2509. doi:10.1016/j.cub.2017.06.075
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of Face Perception. *Annual Review of Neuroscience*, 31(1), 411–437. <http://doi.org/10.1146/annurev.neuro.30.051606.094238>
- Vuong, Q. C., Willenbockel, V., Zimmermann, F. G., Lochy, A., Laguesse, R., Dryden, A., & Rossion, B. (2016). Facelikeness matters: A parametric multipart object set to understand the role of spatial configuration in visual recognition. *Visual Cognition*, 24(7-8), 406-421. doi:10.1080/13506285.2017.1289997

Wardle, S. G., Seymour, K., & Taubert, J. (2017). Characterizing the response to face pareidolia in human category-selective visual cortex. *bioRxiv*, 233387.

Supplementary Tables

Table S1: The number and proportion of images (out of the total no. of 116 images) that had each of the 12 features based on participants rating (standardized score > 0)

Feature	No. of Images	% out of Total
Eyes	76	66%
Mouth	74	64%
Nose	38	33%
Forehead	50	43%
Teeth	33	28%
Ears	24	21%
Eyebrows	36	31%
Hair	36	31%
Frame	66	57%
Proportion	61	53%
Symmetry	59	51%
Emotional Expressiveness	69	59%
Faceness	74	64%

Table S2: The number and proportion of images out of the total no. of 74 images that were perceived as faces (standardized faceness score > 0) that had each of the 12 features based on participants rating (standardized feature score > 0)

Feature	No. of Images	% out of Face Images
Eyes	68	92%
Mouth	68	92%
Nose	34	46%
Forehead	47	64%
Teeth	28	38%
Ears	22	30%
Eyebrows	33	45%
Hair	31	42%
Frame	57	77%
Proportion	59	80%
Symmetry	53	72%
Emotional Expressiveness	67	91%
Faceness	74	100%

Table S3: Regression model which includes the number of features as a predictor together with the 12 features (sorted by t-value).

	B	Std. Error	Beta	t	Sig.
Eyes	0.48	0.07	0.48	6.74	< .001
Mouth	0.25	0.09	0.25	2.88	0.005
Eyebrows	0.11	0.05	0.08	2.03	0.045
Hair	0.08	0.04	0.07	1.83	0.070
Emotional expressiveness	0.12	0.09	0.10	1.28	0.205
<i>Number of features</i>	<i>0.03</i>	<i>0.02</i>	<i>0.13</i>	<i>1.21</i>	<i>0.228</i>
Nose	0.03	0.04	0.03	0.71	0.481
Proportion	0.07	0.13	0.05	0.55	0.586
Forehead	0.01	0.07	0.01	0.13	0.898
Frame	-0.01	0.07	-0.01	-0.16	0.873
Teeth	-0.02	0.04	-0.02	-0.50	0.622
Symmetry	-0.06	0.08	-0.04	-0.76	0.451
Ears	-0.08	0.04	-0.07	-1.78	0.079