



Facial color is an efficient mechanism to visually transmit emotion

Carlos F. Benitez-Quiroz^{a,b}, Ramprakash Srinivasan^{a,b}, and Aleix M. Martinez^{a,b,1}

^aDepartment of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210; and ^bCenter for Cognitive and Brain Sciences, The Ohio State University, Columbus, OH 43210

Edited by Rebecca Saxe, Massachusetts Institute of Technology, Cambridge, MA, and accepted by Editorial Board Member Marlene Behrmann February 16, 2018 (received for review September 12, 2017)

Facial expressions of emotion in humans are believed to be produced by contracting one's facial muscles, generally called action units. However, the surface of the face is also innervated with a large network of blood vessels. Blood flow variations in these vessels yield visible color changes on the face. Here, we study the hypothesis that these visible facial colors allow observers to successfully transmit and visually interpret emotion even in the absence of facial muscle activation. To study this hypothesis, we address the following two questions. Are observable facial colors consistent within and differential between emotion categories and positive vs. negative valence? And does the human visual system use these facial colors to decode emotion from faces? These questions suggest the existence of an important, unexplored mechanism of the production of facial expressions of emotion by a sender and their visual interpretation by an observer. The results of our studies provide evidence in favor of our hypothesis. We show that people successfully decode emotion using these color features, even in the absence of any facial muscle activation. We also demonstrate that this color signal is independent from that provided by facial muscle movements. These results support a revised model of the production and perception of facial expressions of emotion where facial color is an effective mechanism to visually transmit and decode emotion.

face perception | categorization | affect | computer vision

For over 2,300 y (1), philosophers and scientists have argued that people convey emotions through facial expressions by contracting and relaxing their facial muscles (2). The resulting, observable muscle articulations are typically called action units (AUs) (3).

It has been shown that different combinations of AUs yield facial expressions associated with distinct emotion categories (2, 4, 5), although cultural and individual differences exist (6, 7) and context may influence their production and perception (8, 9). These facial expressions are easily interpreted by our visual system, even in noisy environments (10–12).

Here, we test the supplemental hypothesis that facial color successfully transmits emotion independently of facial movements. Our hypothesis is that a face can send emotion information to observers by changing the blood flow or blood composition on the network of blood vessels closest to the surface of the skin (Fig. 1). Consider, for instance, the redness that is sometimes associated with anger or the paleness in fear. These facial color changes can be caused by variations in blood flow, blood pressure, glucose levels, and other changes that occur during emotive experiences (13–15).

Can facial color accurately and robustly communicate emotion categories and valence to observers? That is, when experiencing an emotion, are the skin color changes consistent within expressions of the same emotion category/valence and differential between categories/valences (even across individuals, ethnicity, gender, and skin color)? And are these changes successfully visually interpreted by observers?

The present paper provides supporting evidence for this hypothesis.

First, we demonstrate that specific color features in distinct facial areas are indeed consistent within an emotion category (or valence) but differential between them. Second, we use a machine-learning algorithm to identify the most descriptive color features associated with each emotion category. This allows us to change the color of neutral face images (i.e., faces without any facial muscle movement) to match those of specific emotions. Showing these images to human subjects demonstrates that people do perceive the correct emotion category (and valence) on the face even in the absence of muscle movements. Finally, we demonstrate that the emotion information transmitted by color is additive to that of AUs. That is, the contributions of color and AUs to emotion perception are (at least partially) independent.

These results provide evidence for an efficient and robust communication of emotion through modulations of facial color. These results call for a revised model of the production and visual perception of facial expressions where facial color is successfully used to transmit and visually identify emotion.

Results

Our first goal is to determine whether the production of distinct facial expressions of emotion (plus neutral ones) can be discriminated based on color features alone, having eliminated all shading information. By discriminant, we mean that there are facial color features that are consistent within an emotion category (regardless of identity, gender, ethnicity, and skin color) and

Significance

Emotions correspond to the execution of a number of computations by the central nervous system. Previous research has studied the hypothesis that some of these computations yield visually identifiable facial muscle movements. Here, we study the supplemental hypothesis that some of these computations yield facial blood flow changes unique to the category and valence of each emotion. These blood flow changes are visible as specific facial color patterns to observers, who can then successfully decode the emotion. We present converging computational and behavioral evidence in favor of this hypothesis. Our studies demonstrate that people identify the correct emotion category and valence from these facial colors, even in the absence of any facial muscle movement.

Author contributions: A.M.M., with the help of C.F.B.-Q., designed research; C.F.B.-Q. and R.S. performed research; C.F.B.-Q., R.S., and A.M.M. analyzed data; C.F.B.-Q., R.S., and A.M.M. wrote the paper; and A.M.M. conceived the main idea.

Conflict of interest statement: The Ohio State University has submitted a patent application that includes the recognition of emotion from facial color.

This article is a PNAS Direct Submission. R.S. is a guest editor invited by the Editorial Board.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence should be addressed. Email: martinez.158@osu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716084115/-DCSupplemental.

Published online March 19, 2018.



Fig. 1. Emotions are the execution of a number of computations by the nervous system. Previous research has studied the hypothesis that some of these computations yield muscle articulations unique to each emotion category. However, the human face is also innervated with a large number of blood vessels, as illustrated (veins in blue, arteries in red). These blood vessels are very close to the surface of the skin. This makes variations in blood flow visible as color patterns on the surface of the skin. We hypothesize that the computations executed by the nervous system during an emotion yield color patterns unique to each emotion category. (Copyright The Ohio State University.)

differential between them. We test this in experiment 1. Experiment 2 identifies the color changes coding each emotion category. Experiment 3 incorporates these color patterns to neutral faces to show that people visually recognize emotion based on color features alone. And experiment 4 shows that this emotive color information is independent from that transmitted by AUs.

Experiment 1. We use images of 184 individuals producing 18 facial expressions of emotions (5) (*Materials and Methods, Images*). These include both genders and many ethnicities and races (*SI Appendix, Fig. S1A and Extended Methods, Databases*). Previous work has demonstrated that the pattern of AU activation in these facial expressions is consistent within an emotion category and differential between them (5, 7, 16). But are facial color patterns also consistent within each emotion and differential between emotions?

To answer this question, we built a linear machine-learning classifier using linear discriminant analysis (LDA) (17, 18) (*Materials and Methods, Classification*). First, each face is divided into 126 local regions defined by 87 anatomical landmark points (*SI Appendix, Fig. S2*). These facial landmark points define the contours of the face (jaw and crest line) and internal facial components (brows, eyes, nose, and mouth). A Delaunay triangulation of these points yields the local areas of the face shown in *SI Appendix, Fig. S2*. These local regions correspond to local areas of the network of blood vessels illustrated in Fig. 1.

Let the feature vector \tilde{x}_{ij} define the isoluminant opponent color (19) features of the j th emotion category as produced by the i th individual (*Materials and Methods, Color Space*).

If indeed local facial color is consistent within each emotion category and differential between them, then a linear classifier should discriminate these feature vectors as a function of their emotion category. We use a simple linear model and disjoint training and testing sets to avoid overfitting.

Testing was done using 10-fold cross-validation (*Materials and Methods, Experiment 1*). This means that we randomly split the dataset of all our images into 10 disjoint subsets of equal size. Nine of these subsets are used to train our linear classifier, while the 10th is used for testing, i.e., to determine how accurate this classifier is on unobserved data. This process is repeated 10

times, each time leaving a different subset out for testing. The average classification accuracy and SD on the left-out subsets is computed, yielding the confusion table results shown in Fig. 2A. Columns in this confusion table specify the true emotion category, and rows correspond to the classification given by our classifier. Thus, diagonal entries correspond to correct classifications and off-diagonal elements to confusions.

The correct classification accuracy, averaged over all emotion categories, is 50.15% (chance = 5.5%), and all emotion categories yield results that are statistically significant above chance ($P < 0.001$). A power analysis yielded equally supportive results (*SI Appendix, Table S1*). The high effect size and low P value, combined with the small number of confusions made by this linear classifier, provide our first evidence for the existence of consistent color features within each emotion category that are also differential across emotions.

These results are not dependent on the skin color as demonstrated in *SI Appendix, Fig. S3A and Results Are Not Dependent on Skin Color*.

Classification of positive vs. negative valence by this machine-learning algorithm was equally supportive of our hypothesis. Classification was 92.93% (chance = 50%), $P < 10^{-90}$.

Experiment 2. To identify the most discriminant (diagnostic) color features, we repeated the above computational analysis but using a one-vs.-all approach (*Materials and Methods, Experiment 2*). This means that the images of the target emotion are used as the samples of one class, while the images of all other (nontarget) emotions are assigned to the second class. Specifically, LDA is used to yield the color models \tilde{x}_j of each of the target emotions and the color model of the nontarget emotions \tilde{x}_{-j} , $j = 1, \dots, 18$. That is, \tilde{x}_j and \tilde{x}_{-j} are defined by the most discriminant color features given by LDA.

Using this LDA classifier and a 10-fold cross-validation procedure yields the accuracies and standard errors shown in Fig. 2B. As we can see in Fig. 2B, the classification accuracies of all emotion categories are statistically better than chance ($P < 10^{-50}$, chance = 50%; *SI Appendix, Table S2*). These results are not dependent on skin color as demonstrated in *SI Appendix, Fig. S3B*.

Importantly, the relevance of the color channels in each local face area varies across emotion categories, as expected (*SI Appendix, Fig. S4A*). This result thus supports our hypothesis that experiencing distinct emotions yields differential color patterns on the face. That is, when producing facial expressions of distinct emotions, the observable facial color differs. But when distinct people produce facial expressions of the same emotion, these discriminant color features remain constant. Furthermore, the most discriminant color areas are mostly different from the most discriminant shape changes caused by AUs (*SI Appendix, Fig. S4B*). Thus, color features that transmit emotion are generally independent of AUs.

The identified color features can now be added to images of facial expressions, as shown in Fig. 3A. In addition, the dimensionality of the color space is 18, with each dimension defining an emotion category (*SI Appendix, Fig. S5 and Dimensionality of the Color Space*), further supporting our hypothesis.

Since the production of spontaneous expressions generally differs from that of posed expressions (16, 20), we also assessed the ability of isoluminant color to discriminate spontaneous expressions. Specifically, the DISFA (Denver Intensity of Spontaneous Facial Action) database (21) includes a large number of samples of four emotion categories plus neutral. Using these images and the above machine-learning approach yields an average classification accuracy of 95.53% (chance = 20%; *SI Appendix, Table S3*) (Fig. 2C). We note that spontaneous expressions of distinct emotions are more clearly discriminated than posed expressions, as expected.

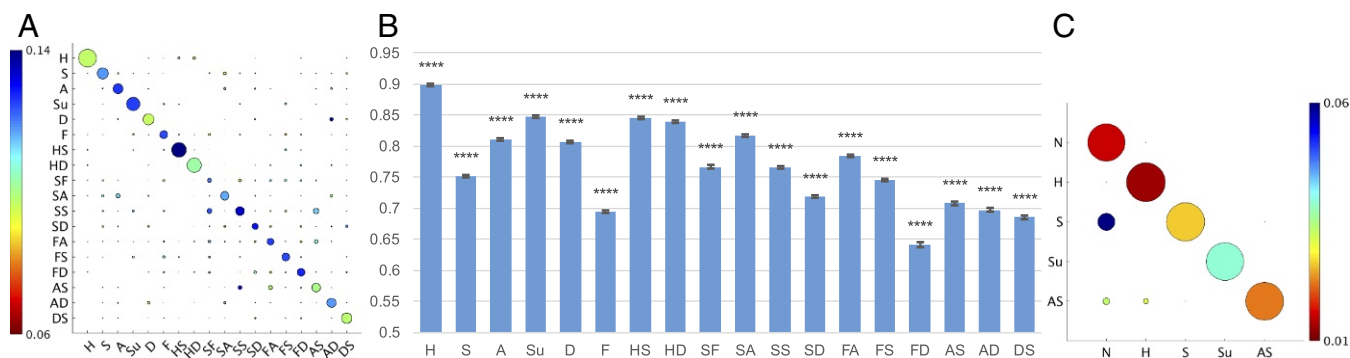


Fig. 2. Consistent with our hypothesis, specific color features remain unchanged for images of the same emotion category but are differential between images of distinct categories. (A) Confusion table of the results of a k -way classification with LDA (experiment 1). Rows indicate the predicted category and columns the true category. The size of the circle specifies the accuracy of recognition (i.e., larger circles equal higher accuracy) and the color the SD across identities. (B) Results of the two-way classifications using LDA (experiment 2). The emotion category indicated in the x axis corresponds to the target emotion. Blue bars indicate accuracy of classification between the target and nontarget emotions, and error bars specify SE. All results are statistically significant (change = 50%); **** $P < 10^{-50}$. (C) Confusion table of a k -way classification on the four emotion categories in DISFA. A, angry; AD, angrily disgusted; AS, angrily surprised; D, disgusted; DS, disgustingly surprised; F, fearful; FA, fearfully angry; FD, fearfully disgusted; FS, fearfully surprised; H, happy; HD, happily disgusted; HS, happily surprised; N, neutral; S, sad; SA, sadly angry; SD, sadly disgusted; SF, sadly fearful; SS, sadly surprised; Su, surprised.

Experiment 3. The color models of experiment 2 are now used to modify the neutral faces of the 184 individuals in our database. That is, we change the color of neutral faces using the discriminant models \tilde{x}_j and \tilde{x}_{-j} identified above.

Specifically, the feature vector \mathbf{x}_{in} associated with image \mathbf{I}_{in} (of a neutral face) is modified using the equation $\mathbf{y}_{ij} = \mathbf{x}_{in} + \alpha(\tilde{x}_j - \tilde{x}_{-j})$, where $\alpha = 1$ (Materials and Methods, Experiment 3). Let us call the image associated with the feature vector \mathbf{y}_{ij} , $\tilde{\mathbf{I}}_{ij}$. We used this approach to add the color of the following six emotion categories to neutral faces: angry, disgusted, happy, happily disgusted, sad, and fearfully surprised (Fig. 3B). This yields a new set of neutral faces, $\tilde{\mathbf{I}}_{ij}$, $j = 1, \dots, 6$.

These images were then shown in pairs to participants (Fig. 3C and SI Appendix, Fig. S6). Each pair had an image corresponding to the target emotion, $\tilde{\mathbf{I}}_{ij}$, and another one corresponding to one of the nontarget emotions, $\tilde{\mathbf{I}}_{ik}$, $k \neq j$. All target, nontarget emotion pairs were tested. Participants were asked to identify which one of the two images in the screen appears to express emotion j , $\tilde{\mathbf{I}}_{ij}$ or $\tilde{\mathbf{I}}_{ik}$. Half of the time $\tilde{\mathbf{I}}_{ij}$ was the left image and half of

the time it was the right one. This order was randomized across participants. Participants responded by keypress.

The percentage of times participants selected $\tilde{\mathbf{I}}_{ij}$ over $\tilde{\mathbf{I}}_{ik}$ is given in Fig. 4A, with $j = \{1, \dots, 6\}$. As can be seen in Fig. 4A, participants consistently preferred the neutral face with the target emotion color, $\tilde{\mathbf{I}}_{ij}$, rather than the image with the colors of a nontarget emotion, $\tilde{\mathbf{I}}_{ik}$. These results were statistically significant, $P < 0.01$ (SI Appendix, Table S4). The recognition of positive vs. negative valence was also significant; accuracy = 82.65% (chance = 50%), $P < 10^{-4}$.

Next, we tested the ability of subjects to classify each $\tilde{\mathbf{I}}_{ij}$ image into one of six possible choices, i.e., a six-alternative forced-choice experiment (Materials and Methods, Six-alternative forced-choice experiment) (Fig. 3D). The results are shown in Fig. 4B; average classification = 32.92% (chance = 16.67%). All results are statistically significant, $P < 0.01$, except for happily disgusted, which is confused with happy. Nonetheless, this confusion yields the correct perception of valence. In fact, average accuracy of the recognition of valence in this challenging

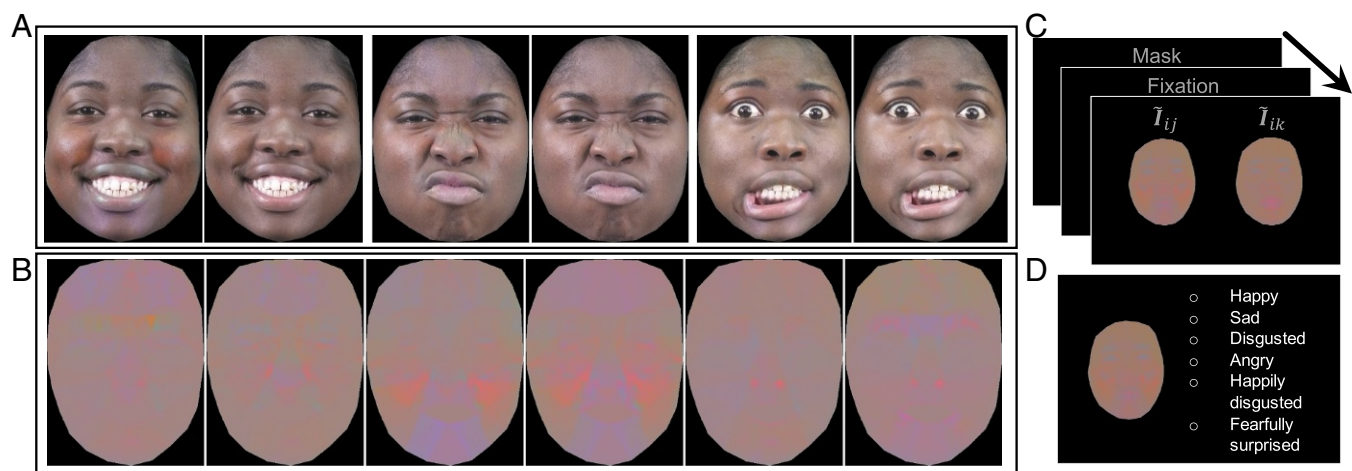


Fig. 3. (A) The first pair of images corresponds to the expression of happiness, the second pair to expression of anger, and the last pair to expression of surprise. For each pair of images, the Left image includes the diagnostic facial colors identified in experiment 2 (SI Appendix, Fig. S7). (B) Sample images $\tilde{\mathbf{I}}_{ik}$ of experiment 3. Left to Right: angry, disgusted, happy, happily disgusted, sad, fearfully surprised. (C and D) Sample trials of the two- and six-alternative forced-choice method of experiment 3.

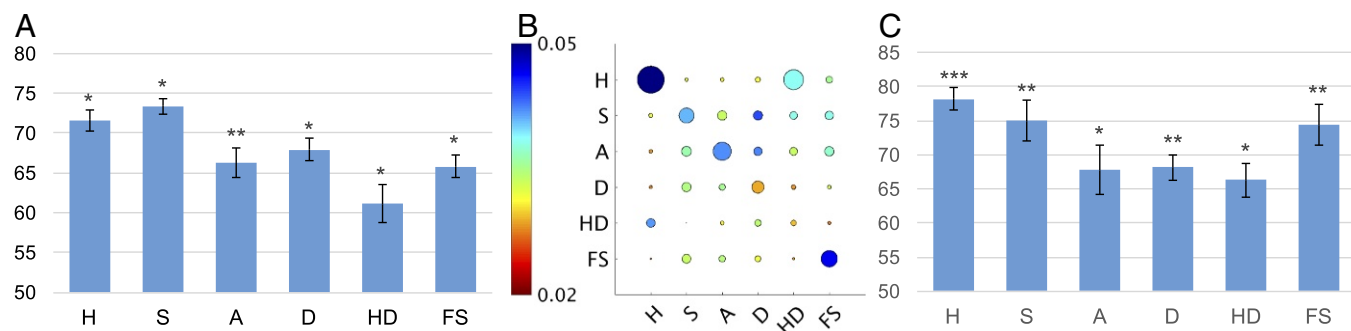


Fig. 4. Results of experiments 3 and 4. (A) The y axis shows the proportion of times participants indicated that the neutral image with the diagnostic color features of the target emotion specified in the x axis (I_{ij}) expresses that emotion better than the neutral face with the colors of a nontarget emotion (I_{ik}). Images are in isoluminant color space; i.e., images do not have any shading information. (B) Confusion table of a six-alternative forced choice experiment. Subjects are shown a single image at a time and asked to classify it into a set of six emotion categories. The size of the circles specifies classification accuracies, and colors are SD. Accuracies are better than chance except for happily disgusted, which is confused for happy. (C) The y axis shows the proportion of times participants indicated that a face with the AUs and diagnostic color features of the target emotion specified in the x axis (I_{ij}) expresses that emotion better than a face with the AUs of the target emotion and the colors of a nontarget emotion (I_{ik}). *** $P < 0.0001$, ** $P < 0.001$, * $P < 0.01$.

experiment is 72.9% (chance = 50%, $P < 10^{-5}$) (SI Appendix, Table S5).

Experiment 4. One may wonder whether the emotion perception effect demonstrated in the previous experiment disappears when the facial expression also includes AUs. That is, are the emotion information visually transmitted by facial muscle movements and diagnostic color features independent from one another? If the color features do indeed transmit (at least some) independent information from that of AUs, then having the color of the target emotion j should yield a clearer perception of the target emotion than having the colors of a nontarget emotion k ; i.e., congruent colors make categorization easier, and incongruent colors make it more difficult.

To test this prediction, we created two additional image sets. In the first image set, we added the color model of emotion category j to the images of the 184 individuals expressing the same emotion with AUs. Formally, image I_{ij}^+ corresponds to the modified feature vector $\mathbf{z}_{ij}^+ = \mathbf{x}_{ij} + \alpha(\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_{-j})$. Here \mathbf{x}_{ij} is the feature vector of the facial expression of emotion j as expressed by individual i (i.e., the feature vector of image I_{ij}), and $\alpha = 1$ (Materials and Methods, Experiment 4; Fig. 3A; and SI Appendix, Fig. S7). In the second set, we added the color features of a nontarget emotion k in an image of a facial expression of emotion j , $k \neq j$. The resulting images, I_{ij}^k are given by the feature vectors $\mathbf{z}_{ij}^k = \mathbf{x}_{ij} + \alpha(\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_{-k})$, $\alpha = 1$ (SI Appendix, Fig. S7). Thus, I_{ij}^+ is an image with the facial color hypothesized to transmit emotion j , while I_{ij}^k is an image with the diagnostic colors of emotion k instead, $k \neq j$.

Participants completed a two-alternative forced-choice experiment where they had to indicate which of two images of facial expressions convey emotion j more clearly, I_{ij}^+ or I_{ij}^k , $k \neq j$ (Materials and Methods, Experiment 4 and SI Appendix, Fig. S7). The location of I_{ij}^+ and I_{ij}^k (left or right) was randomized. The proportion of times subjects selected I_{ij}^+ as more clearly expressive of emotion j than I_{ij}^k is shown in Fig. 4C. As predicted, the selection of I_{ij}^+ was significantly above chance (chance = 50%). These results are statistically significant, $P < 0.01$ (SI Appendix, Table S6).

The recognition of valence was even stronger; average classification = 85.01% (chance = 50%, $P < 10^{-5}$).

These results support our hypothesis of an (at least partially) independent transmission of emotion signal by color features.

Discussion

Emotions are the execution of a number of computations by the nervous system. For over two millennia, studies have evaluated the hypothesis that some of these computations yield visible facial muscle changes (called AUs) specific to each emotion (1, 2, 4, 5, 7, 22).

The present paper studied the supplemental hypothesis that these computations result in visible facial color changes caused by emotion-dependent variations in facial blood flow. The studies reported in the present paper provide evidence favoring this hypothesis. Crucially, this visual signal is correctly interpreted even in the absence of facial muscle movement. Additionally, the emotion signal transmitted by color is additive to that encoded in the facial muscle movements. Thus, the emotion information transmitted by color is at least partially independent from that by facial movements.

Supporting our results, the human retina (23) and later visual areas (24) have cells specialized for different types of stimuli. Some cells are tuned to detect motion, such as those caused by facial movements. Other cells are specialized for color perception. The visual analysis of emotion using these two separate systems (motion and color) adds robustness to the emotion signal, allowing people to more readily interpret it in noisy environments. Moreover, the existence of two separate neural mechanisms for the analysis of motion and color provides a plausible explanation for the independence of emotion information transmitted by facial movements and facial color.

Also in support of our results is the observation that human faces are mostly bare, whereas those of other primates are covered by facial hair (25). This bareness facilitates the transmission of emotion through the modulation of facial color in humans. This could mean that the transmission of an emotion signal through facial color is a mechanism not available to all primates and may be a result of recent evolutionary forces.

These findings call for a revisit to emotion models. For instance, the present paper has demonstrated that color can successfully communicate 18 distinct emotion categories to an observer as well as positive and negative valence, but more categories or degrees of valences are likely (7, 26). For example, it is possible that new combinations of AUs and colors, not tested in the present study, can yield previously unidentified facial expressions of emotion. This is a fundamental question in emotion research likely to influence the definition of emotion and the role of emotions in high-level cognitive tasks (26, 27).

The results of the present study also call for the study of the brain pathways associated with emotion perception using color

alone and in conjunction with facial movements. Previous studies have suggested the existence of a variety of pathways in the interpretation of facial expressions of emotion, but the specificity of some of these pathways is poorly understood (28–30). Also, understanding independent domain-specific mechanisms such as the ones supported by our results is a topic of high value in neuroscience (31, 32).

Current computational models of the perception of facial expressions of emotion (7) will also need to be extended or revised to include the contribution of facial color and color perception. Similarly, algorithms in artificial intelligence (e.g., computer vision and human–robot interaction) will need to be able to interpret emotion through these two independent mechanisms if they are to be indistinguishable from humans, i.e., to pass a visual Turing test (33).

Also consistent with our hypothesis, studies in computer graphics demonstrate the importance that color plays in the expression of emotion (34–37).

Our findings are also significant in the study of psychopathologies. A recurring characteristic in psychopathologies is an atypical perception of facial expressions of emotion (38). But is this caused by limitations in interpreting facial movements or color? Or is this dependent on the disorder?

Finally, the results of the present study support the view that the perception of color can have an emotional interpretation, as painters have been exploiting for years (39). Of note are the paintings of Mark Rothko, consisting of blurred blocks of colors combined to yield the perception of emotion. The studies reported in the present paper could help understand the perception of emotion so eloquently conveyed by this and other painters.

Materials and Methods

Approval and Consent. The experiment design was approved by the Office of Responsible Research Practices at The Ohio State University. Subjects provided written consent.

Images. We used images of 184 individuals expressing 18 emotion categories, plus neutral. The images are from refs. 5 and 21. These images have been extensively validated for consistency of production and recognition (7). The emotion categories are happy, sad, angry, disgusted, surprised, fearful, happily surprised, happily disgusted, sadly fearful, sadly angry, sadly surprised, sadly disgusted, fearfully angry, fearfully surprised, fearfully disgusted, angrily surprised, angrily disgusted, and disgustedly surprised. The dataset also includes the neutral face (*SI Appendix, Extended Methods, Databases*).

Areas of the Face. Denote each face color image of $p \times q$ pixels as $I_{ij} \in \mathbb{R}^{p \times q \times 3}$ and the r landmark points of each of the facial components of the face as $\mathbf{s}_{ij} = (\mathbf{s}_{ij1}, \dots, \mathbf{s}_{ijr})$, $\mathbf{s}_{ijk} \in \mathbb{R}^2$ the 2D coordinates of a landmark point on the image. Here, i specifies the subject and j the emotion category and we used $r = 66$ (*SI Appendix, Fig. S2*). Delaunay triangulation (40) is used to create the triangular local areas shown in *SI Appendix, Fig. S2*. This triangulation yields a total of 148 local areas. Six of these triangles define the interior of the mouth and 16 the inside of the eyes (sclera and iris). Since blood changes in these areas are not generally visible, they are removed from further consideration, leaving a total of 126 triangular local areas. Let $D = \{d_1, \dots, d_{126}\}$ be a set of functions that return the pixels of each of these local regions; i.e., $d_k(I_{ij})$ is a vector including the l pixels within the k th Delaunay triangle in image I_{ij} ; i.e., $d_k(I_{ij}) = (d_{ijk1}, \dots, d_{ijk40})^T$, where $\mathbf{d}_{ijks} = (d_{ijks1}, d_{ijks2}, d_{ijks3})^T \in \mathbb{R}^3$ defines the values of the LMS channels of each pixel.

Color Space. The above-defined feature vectors are mapped onto the isoluminant color space given by the opponent color channels yellow–blue and red–green. Formally $\mathbf{h}_{ijks} = (d_{ijks1} + d_{ijks2} - d_{ijks3}, d_{ijks1} - d_{ijks2})^T \in \mathbb{R}^2$. We compute the first and second moments (i.e., mean and variance) of the data in this color space; i.e., $\mu_{ijk} = I^{-1} \sum_{s=1}^l \mathbf{h}_{ijks}$, $\sigma_{ijk} = I^{-1} \sum_{s=1}^l \sqrt{(\mathbf{h}_{ijks} - \mu_{ijk})^2}$. Every image I_{ij} is now using the following feature vector of color statistics, $\mathbf{x}_{ij} = (\mu_{ij1}^T, \sigma_{ij1}^T, \dots, \mu_{ij126}^T, \sigma_{ij126}^T)^T \in \mathbb{R}^{504}$. Using the same modeling, we

define the color feature vector of every neutral face as $\mathbf{x}_{in} = (\mu_{in1}^T, \sigma_{in1}^T, \dots, \mu_{in126}^T, \sigma_{in126}^T)^T$. We also use the normalization proposed by Bratko et al. (41), which maintains the properties of saturation and hue in color space. This yields the vectors $\hat{\mathbf{x}}_{ij}$ and $\hat{\mathbf{x}}_{in}$, respectively.

Classification. LDA is computed on the above-defined color space. Formally, the color space is defined by the eigenvectors associated with nonzero eigenvalues of the matrix $\Sigma_x^{-1} \mathbf{S}_B$ (17), where $\Sigma_x = \sum_{i=1}^m \sum_{j=1}^C (\hat{\mathbf{x}}_{ij} - \mu)(\hat{\mathbf{x}}_{ij} - \mu)^T + \delta \mathbf{I}$ is the (regularized) covariance matrix, $\mathbf{S}_B = \sum_{j=1}^C (\bar{\mathbf{x}}_j - \mu)(\bar{\mathbf{x}}_j - \mu)^T$ is the between-class scatter matrix, $\bar{\mathbf{x}}_j = m^{-1} \sum_{i=1}^m \hat{\mathbf{x}}_{ij}$ are the class means, $\mu = (Cm)^{-1} \sum_{i=1}^m \sum_{j=1}^C \hat{\mathbf{x}}_{ij}$, \mathbf{I} is the identity matrix, $\delta = 0.01$, and C is the number of classes.

Experiment 1. The LDA isoluminant color space is computed with $C = 18$ and $C = 5$ when using the datasets of refs. 5 and 21, respectively; i.e., each emotion and the neutral expression are a class. For 10-fold cross-validation, we divide the samples into 10 disjoint subsets $S = \{S_1, \dots, S_{10}\}$, each subset S_t having the same number of samples. This division is done in a way that the number of samples in each emotion category (plus neutral) is equal in every subset. To compute the results reported in experiment 1, we repeat the following procedure with $t = 1, \dots, 10$: All of the subsets, except S_t , are used to compute Σ_x and \mathbf{S}_B . The samples in subset S_t , which were not used to compute the LDA subspace \mathcal{F} , are projected onto \mathcal{F} . Each of the test sample feature vectors $\mathbf{t}_r \in S_t$ is assigned to the emotion category of the nearest category mean, given by the Euclidean distance, $e_r^* = \arg\min_e (\mathbf{t}_r - \bar{\mathbf{x}}_e)^T (\mathbf{t}_r - \bar{\mathbf{x}}_e)$, in \mathcal{F} , and $\bar{\mathbf{x}}$ is the mean feature vector of all feature vectors in S_e . The classification accuracy over all of the test samples $\mathbf{t}_r \in S_t$ is given by $\mu_{t\text{-fold}} = n_t^{-1} \sum_{\mathbf{t}_r \in S_t} \mathbf{1}_{\{e_r^* = y(\mathbf{t}_r)\}}$, where n_t is the number of samples in S_t , $y(\mathbf{t}_r)$ is the oracle function that returns the true emotion category of sample \mathbf{t}_r , and $\mathbf{1}_{\{e_r^* = y(\mathbf{t}_r)\}}$ is the zero–one loss which equals one when $e_r^* = y(\mathbf{t}_r)$ and zero elsewhere. Since $t = 1, \dots, 10$, we repeat this procedure 10 times and compute the mean classification accuracy $\mu_{10\text{-fold}} = 0.1 \sum_{t=1}^{10} \mu_{t\text{-fold}}$. The SD of the cross-validated classification accuracies is $\sigma_{10\text{-fold}} = 0.1 \sum_{t=1}^{10} (\mu_{t\text{-fold}} - \mu_{10\text{-fold}})^2$.

Experiment 2. We identify the most discriminant color features of each emotion category by repeating the approach described above (*Materials and Methods, Experiment 1*) 18 times, each time assigning the samples of one emotion category to class 1 (i.e., the target emotion) and the samples of all other emotion categories to class 2 (i.e., the nontarget emotions). Thus, $C = 2$. Formally, $S_c = \{\mathbf{t}_r | y(\mathbf{t}_r) = c\}$ and $S_{-c} = \{\mathbf{t}_r | y(\mathbf{t}_r) \neq c\}$, with $c = \{1, \dots, 18\}$. We use the same 10-fold cross-validation procedure and nearest-mean classifier described in experiment 1. To avoid biases due to the sample imbalance, we apply downsampling on S_{-c} . Specifically, we repeated this procedure 1,000 times, each time drawing a random sample from S_{-c} to match the number of samples in S_c . Then, we compute the average. Let $\Sigma_x^{-1} \mathbf{S}_B \mathbf{v}_1 = \mathbf{v}_1 \lambda_1$. This yields the discriminant vector $\mathbf{v}_1 = (v_{1,1}, \dots, v_{1,504})^T$, with $\lambda_1 > 0$. The color model of emotion j is thus given by $\hat{\mathbf{x}}_j = \mathbf{v}_1^T \hat{\mathbf{x}}_j$. Similarly, $\hat{\mathbf{x}}_{-j} = \mathbf{v}_1^T \bar{\mathbf{x}}_{-j}$, where $\bar{\mathbf{x}}_{-j}$ is the mean feature vector of all feature vectors not in S_j (i.e., in S_{-j}). The reprojection of these vectors into the original space is called reconstruction. It is given by $\tilde{\mathbf{x}}_j = \hat{\mathbf{x}}_j \mathbf{v}_1^T$ and $\tilde{\mathbf{x}}_{-j} = \hat{\mathbf{x}}_{-j} \mathbf{v}_1^T$.

Experiment 3.

Subjects. Twenty human subjects (7 women; mean age, 23.92 y) with normal or corrected to normal vision participated in this experiment. Subjects were tested for color blindness.

Stimuli. We modify the original neutral faces, I_{in} , $i = 1, \dots, 184$. The resulting images \tilde{I}_{ij} are given by the color feature vectors $\mathbf{y}_{ij} = \hat{\mathbf{x}}_{in} + \alpha(\hat{\mathbf{x}}_j - \hat{\mathbf{x}}_{-j})$, where $\alpha = 1$, and $\hat{\mathbf{x}}_{in}$ is the feature vector with the average color and luminance of all of the pixels of \mathbf{x}_{in} . This equation adds the diagnostic color features of emotion j to a neutral expression. The resulting images have the same luminance in every pixel (Fig. 3B). We selected the six emotion categories shown in Fig. 3B to keep the behavioral experiment shorter than 35 min.

Design and procedure. Subjects were comfortably seated 50 cm from a 21-inch cathode ray tube (CRT) monitor. We used a block design. In each of the six blocks, we tested one of the six emotion categories. At the beginning of each block, a screen indicated the target emotion, followed by a screen with a sample image of that emotion, i.e., an image of a facial expression of emotion j with the prototypical AUs present. Each of the six blocks consists of 20 trials. Each trial is a two-alternative forced-choice behavioral experiment. Fig. 3C and *SI Appendix, Fig. S5* illustrate a typical time line of a trial. Participants first viewed a blank screen for 500 ms, followed by a white

fixation cross for 500 ms, followed by the image pair \tilde{I}_{ij} and \tilde{I}_{ik} , $k \neq j$ (Fig. 3C). Responses were recorded by keypress. Statistical significance is given by a one-tailed *t* test.

Six-alternative forced-choice experiment. In the six-alternative forced choice experiment (6AFC), participants saw a single image in each trial, and there was a single block with all images randomized. Participants chose one of six possible emotion category labels provided to the right of the image (Fig. 3D). Selection was done with the computer mouse.

Experiment 4.

Subjects. Twenty human subjects (8 women; mean age, 24.45 y) participated. All subjects passed a color blindness test.

Stimuli. We modified the images of six facial expressions of emotion with the prototypical AUs active, I_{ij} , $i = 1, \dots, 184$, $j = 1, \dots, 6$ (Fig. 3A). We modified these images to have the colors of a target or nontarget emotion. The images with the target emotion colors added are I_{ij}^+ and are given by the modified feature vector $\mathbf{z}_{ij} = \mathbf{x}_{ij} + \alpha(\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_{-j})$, $i = 1, \dots, 184$, $\alpha = 1$. Hence, I_{ij}^+ are images of facial expressions with the AUs of emotion category j plus the colors of emotion j (Fig. 3 and *SI Appendix, Fig. S7*). The images with

the modified color of nontarget emotions I_{ij}^- are given by the feature vector $\mathbf{z}_{ij} = \mathbf{x}_{ij} + \alpha(\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_{-k})$ with $k \neq j$, $\alpha = 1$ (*SI Appendix, Fig. S7*).

Design and procedure. Participants were comfortably seated 50 cm from a 21-inch CRT monitor. We used the same block design of experiment 3. Each trial is a two-alternative forced choice behavioral experiment. *SI Appendix, Fig. S8* illustrates a typical timeline of a trial. In each trial, participants viewed a blank screen for 500 ms, followed by a white fixation cross for 500 ms, followed by the image pair I_{ij}^+ , I_{ik}^- . Participants are instructed to indicate whether the left or right image expresses emotion j more clearly. Responses are recorded by keypress. The positions (left, right) of I_{ij}^+ and I_{ik}^- are randomized. Their presentation order (trials) is also randomized across participants. Statistical significance is given by a one-tailed *t* test, which yielded the values 0.0000081 (H), 0.0005 (S), 0.0051 (A), 0.00015 (D), 0.0016 (HD), and 0.0006 (FS) (abbreviations H, S, A, D, HD, and FS are defined in Fig. 2).

ACKNOWLEDGMENTS. We thank the reviewers and editor for constructive feedback. We also thank Ralph Adolphs and Lisa Feldman Barrett for comments and Delwin Lindsey and Angela Brown for pointing us to ref. 25. This research was supported by the National Institutes of Health, Grant R01-DC-014498, and the Human Frontier Science Program, Grant RGP0036/2016.

- Aristotle (1939) *Aristotle: Minor Works* (Harvard Univ Press, Cambridge, MA).
- Duchenne GB (1990) *The Mechanism of Human Facial Expression* (Cambridge Univ Press, Paris).
- Ekman P, Rosenberg EL (1997) *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (Oxford Univ Press, New York).
- Ekman P, Sorenson ER, Friesen WV (1969) Pan-cultural elements in facial displays of emotion. *Science* 164:86–88.
- Du S, Tao Y, Martinez AM (2014) Compound facial expressions of emotion. *Proc Natl Acad Sci USA* 111:E1454–E1462.
- Jack RE, Garrod OG, Yu H, Caldara R, Schyns PG (2012) Facial expressions of emotion are not culturally universal. *Proc Natl Acad Sci USA* 109:7241–7244.
- Martinez AM (2017) Visual perception of facial expressions of emotion. *Curr Opin Psychol* 17:27–33.
- Barrett LF, Kensinger EA (2010) Context is routinely encoded during emotion perception. *Psychol Sci* 21:595–599.
- Etkin A, Büchel C, Gross JJ (2015) The neural bases of emotion regulation. *Nat Rev Neurosci* 16:693–700.
- Smith FW, Schyns PG (2009) Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychol Sci* 20:1202–1208.
- Du S, Martinez AM (2011) The resolution of facial expressions of emotion. *J Vis* 11:24.
- Srinivasan R, Golomb JD, Martinez AM (2016) A neural basis of facial action recognition in humans. *J Neurosci* 36:4434–4442.
- Blake TM, Varnhagen CK, Parent MB (2001) Emotionally arousing pictures increase blood glucose levels and enhance recall. *Neurobiol Learn Mem* 75:262–273.
- Cahill L, Prins B, Weber M, McGaugh JL (1994) β -adrenergic activation and memory for emotional events. *Nature* 371:702–704.
- Garfinkel SN, Critchley HD (2016) Threat and the body: How the heart supports fear processing. *Trends Cognit Sci* 20:34–46.
- Fabian Benitez-Quiroz C, Srinivasan R, Martinez AM (2016) Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. *Proc IEEE Conf Comput Vis Pattern Recognit*, 10.1109/CVPR.2016.600.
- Martinez AM, Zhu M (2005) Where are linear feature extraction methods applicable? *IEEE Trans Pattern Anal Mach Intell* 27:1934–1944.
- Martinez AM, Kak AC (2001) PCA versus LDA. *IEEE Trans pattern Anal Mach Intell* 23:228–233.
- Gegenfurtner KR (2003) Cortical mechanisms of colour vision. *Nat Rev Neurosci* 4:563–572.
- Fernández-Dols JM, Crivelli C (2013) Emotion and expression: Naturalistic studies. *Emot Rev* 5:24–29.
- Mavadati SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF (2013) DISFA: A spontaneous facial action intensity database. *IEEE Trans Affect Comput* 4:151–160.
- Hjortsjö CH (1969) *Man's Face and Mimic Language* (Student Litteratur, Lund, Sweden).
- Euler T, Haverkamp S, Schubert T, Baden T (2014) Retinal bipolar cells: Elementary building blocks of vision. *Nat Rev Neurosci* 15:507–519.
- Bach M, Hoffmann MB (2000) Visual motion detection in man is governed by non-retinal mechanisms. *Vis Res* 40:2379–2385.
- Changizi MA, Zhang Q, Shimojo S (2006) Bare skin, blood and the evolution of primate colour vision. *Biol Lett* 2:217–221.
- Spunt RP, Adolphs R (2017) The neuroscience of understanding the emotions of others. *Neurosci Lett*, in press.
- Barrett LF (2017) Categories and their role in the science of emotion. *Psychol Inq* 28:20–26.
- Méndez-Bértolo C, et al. (2016) A fast pathway for fear in human amygdala. *Nat Neurosci* 19:1041–1049.
- Wang S, et al. (2014) Neurons in the human amygdala selective for perceived emotion. *Proc Natl Acad Sci USA* 111:E3110–E3119.
- Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF (2012) The brain basis of emotion: A meta-analytic review. *Behav Brain Sci* 35:121–143.
- Spunt RP, Adolphs R (2017) A new look at domain specificity: Insights from social neuroscience. *Nat Rev Neurosci* 18:559–567.
- Kanwisher N (2010) Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc Natl Acad Sci USA* 107:11163–11170.
- Geman D, Geman S, Hallonquist N, Younes L (2015) Visual Turing test for computer vision systems. *Proc Natl Acad Sci USA* 112:3618–3623.
- Mathur MB, Reichling DB (2016) Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition* 146:22–32.
- Jimenez J, et al. (2010) A practical appearance model for dynamic facial color. *ACM Trans Graphics* 29:141.
- Alkawaz MH, Basori AH (2012) The effect of emotional colour on creating realistic expression of avatar. *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (ACM, New York), pp 143–152.
- de Melo CM, Gratch J (2009) Expression of emotions using wrinkles, blushing, sweating and tears. *International Workshop on Intelligent Virtual Agents*, eds Ruttkay Z, Kipp M, Nijholt A, Vilhjálmsson HH (Springer, Amsterdam), pp 188–200.
- Bruyer R (2014) *The Neuropsychology of Face Perception and Facial Expression* (Psychology Press, New York).
- Kandel ER (2016) *Reductionism in Art and Brain Science: Bridging the Two Cultures* (Columbia Univ Press, New York).
- Delaunay B (1934) Sur la sphere vide. *Bull Acad Sciences USSR VII: Class Sci Math*, pp 793–800.
- Bratkova M, Boulous S, Shirley P (2009) oRGB: A practical opponent color space for computer graphics. *IEEE Comput Graphics Appl* 29:42–55.