

The contribution of color information to rapid face categorization in natural scenes

Charles C.-F. Or

Division of Psychology, School of Social Sciences,
Nanyang Technological University, Singapore
Psychological Sciences Research Institute & Institute of
Neuroscience, University of Louvain, Belgium



Talia L. Retter

Psychological Sciences Research Institute & Institute of
Neuroscience, University of Louvain, Belgium
Department of Psychology, Center for Integrative
Neuroscience, University of Nevada, Reno, USA



Bruno Rossion

Psychological Sciences Research Institute & Institute of
Neuroscience, University of Louvain, Belgium
Université de Lorraine, CNRS, CRAN, F-54000,
Nancy, France
CHRU-Nancy, Service de Neurologie, F-54000,
Nancy, France



Color's contribution to rapid categorization of natural images is debated. We examine its effect on high-level face categorization responses using fast periodic visual stimulation (Rossion et al., 2015). A high-density electroencephalogram (EEG) was recorded during presentation of sequences of natural object images every 83 ms (i.e., at $F = 12.0$ Hz). Natural face images were embedded in the sequence at a fixed interval of $F/9$ (1.33 Hz). There were four conditions: (a) full-color images; (b) grayscale images; and (c) and (d) phase-scrambled images from Conditions 1 and 2, respectively, making faces and objects unrecognizable. Observers' task was to respond to color changes of the fixation cross (Experiment 1). We found face-categorization responses at 1.33 Hz and its harmonics (2.67 Hz, etc.) over occipitotemporal areas, with right-hemisphere dominance; responses to color images were not significantly different from those to grayscale images. Behavioral analysis revealed longer response times when images contained color, despite nearly-all-correct performance in all conditions, suggesting that color change in the task might detract from color's contribution to face categorization. We subsequently changed the task to responding to fixation shape changes so that such response-time differences were eliminated (Experiment 2). The aggregate face-categorization response became 21.6% stronger to color than to grayscale images. This color advantage occurred late, at 290–415 ms after stimulus onset. Our results

suggest that the color advantage for face categorization interacts with behavior, and that color only has a moderate and relatively late contribution to rapid face categorization in natural images.

Introduction

The ability to rapidly categorize a visual stimulus as a face is important for social interaction. Face categorization in natural scenes involves segmentation of faces from the background and discrimination of faces from other nonface objects in the environment (e.g., birds, cars, houses, etc.). Face categorization also implies generalization across superficial differences in visual appearance of faces due to variations in both their intrinsic qualities (e.g., identity, sex, race) and extrinsic environmental factors (e.g., differences in viewpoint, lighting, scale; Rossion, Torfs, Jacques, & Liu-Shuang, 2015). Here, we ask what effect color may have on this face categorization process.

Color is a candidate for contributing to human face categorization at all levels. That is, color may facilitate the segmentation of faces from the background, assist in the discrimination of faces from nonface objects, and/or contribute to the generalization of variant face exemplars, perhaps in conjunction with shape infor-

Citation: Or, C. C.-F., Retter, T. L., & Rossion, B. (2019). The contribution of color information to rapid face categorization in natural scenes. *Journal of Vision*, 19(5):20, 1–20, <https://doi.org/10.1167/19.5.20>.

<https://doi.org/10.1167/19.5.20>

Received November 2, 2017; published May 21, 2019

ISSN 1534-7362 Copyright 2019 The Authors



mation. Importantly, faces share a range of diagnostic color information: under natural lighting conditions, face skin colors vary mainly in intensity but differ little in chromaticity, even across human “races” (Yang & Waibel, 1996); for example, faces never appear green or blue. Indeed, the evolution of human color perception may have been influenced by the use of color to categorize conspecifics (as well as for foraging, etc.) in old world monkeys, who share in human trichromacy (Mollon, 1989). In the present day, the potential diagnosticity of facial color information has inspired computer scientists to use face skin colors for machine face detection (Graf, Chen, Petajan, & Cosatto, 1995; Graf, Cosatto, Gibbon, Kocheisen, & Petajan, 1996; Wu, Chen, & Yachida, 1999; Yang & Waibel, 1996), resulting in superior detection speed compared to other methods (De Dios, 2007).

Previous studies have suggested that diagnostic color plays a role in behavioral tasks measuring some aspects of face categorization, despite disagreements across studies on the precise nature of color’s role. For example, Lewis and Edmonds (2003, 2005) found that, in manual response tasks, the time to detect a face in a scrambled natural scene was shorter with diagnostic color in the scene than with a grayscale, or hue-reversed, display, although diagnostic color information was not necessary to make a face pop out. In a saccadic choice task, Boucart et al. (2016) found that colored faces presented in the visual periphery were categorized more accurately, but not significantly faster, than grayscale faces. In a visual search task for faces in natural scenes (using manual responses), Bindemann and Burton (2009) suggested that a color advantage was restricted to the presence of diagnostic color in the entire face image, as they found that performance (both response time and accuracy) was worse when detecting faces of which only half (either left or right) was in color and the other half in grayscale, than when detecting full-color faces. Bindemann and Burton concluded that simply presenting color on half of the face could not improve face detection, while color information was only useful when tied to the general shape of the face, suggesting combined color and shape processing during face detection.

At the level of face categorization, comparison with other nonface objects might alter color’s role, as color’s potential influence on generalization across face color variations (e.g., in skin color) might interact with color as a diagnostic cue to discriminate between the segmented objects. The potential contribution of color to discriminating faces from nonface objects, as well as generalizing across variable face exemplars, has not been explicitly tested to our knowledge. However, some insights can be brought from behavioral studies on rapid object and scene categorization more generally.

Many studies have disputed whether color information is an important cue for the rapid differentiation between briefly presented natural images. In particular, Delorme, Richard, and Fabre-Thorpe (2000) suggested that the presence of color cues had only weak effects of higher accuracy for categorizing animals and shorter response times for categorizing food in natural scenes in manual tasks. However, the contribution of color has been advocated for by Oliva and Schyns (2000; see also Goffaux et al., 2005), who found better performance (both response time and accuracy) from verbal or manual responses in naming and verification tasks for natural scenes in their natural color (e.g., desert, forest, coastline) than grayscale scenes. Castelhana and Henderson (2008) also found that color produced an advantage for manual behavioral responses in determining whether embedded objects were consistent with the natural scenes, though the majority of these were man-made (e.g., city landscapes). Yao and Einhäuser (2008) reported higher accuracy for cross-species animal categorization when images were presented in color. Overall, however, the uncertainty of color’s role in categorization suggests that color’s contribution is limited—for example, only when the attention demand of the task is high (Yao & Einhäuser, 2008, though see Otsuka & Kawaguchi, 2009).

One way to identify the nature of the contribution of color to face categorization is to investigate categorization responses at a neural level. To our knowledge, no studies have compared neural face categorization responses to color and grayscale images. In a previously mentioned study on scene categorization, Goffaux et al. (2005) reported larger and earlier event-related potentials (ERPs), starting from approximately 150 ms poststimulus onset over frontal channels, for naturally colored scenes. Zhu, Drewes, and Gegenfurtner (2013) also reported larger ERP amplitudes and shorter latencies in P1 and N1 responses, peaking over frontal channels, for color than grayscale images. These results have been interpreted in light of color playing a role in the categorization and memory of images (Goffaux et al., 2005), or as a result of color bringing enhanced attention to the images (Zhu et al., 2013, though note that color brought a behavioral advantage in accuracy but disadvantage in response times in that study, with task-dependent effects).

Thus, at this state of our knowledge, an objective identification and quantification of the contribution of color information to human face categorization (i.e., to specific responses to faces) is still lacking. Here, we attempt to answer whether, and if so, how much, color confers an advantage for rapid (i.e., at a single glance) visual categorization of stimuli as faces in natural images. To this extent, we employ fast periodic visual stimulation (FPVS) coupled with a scalp electroencephalogram (EEG), an approach that provides an objective,

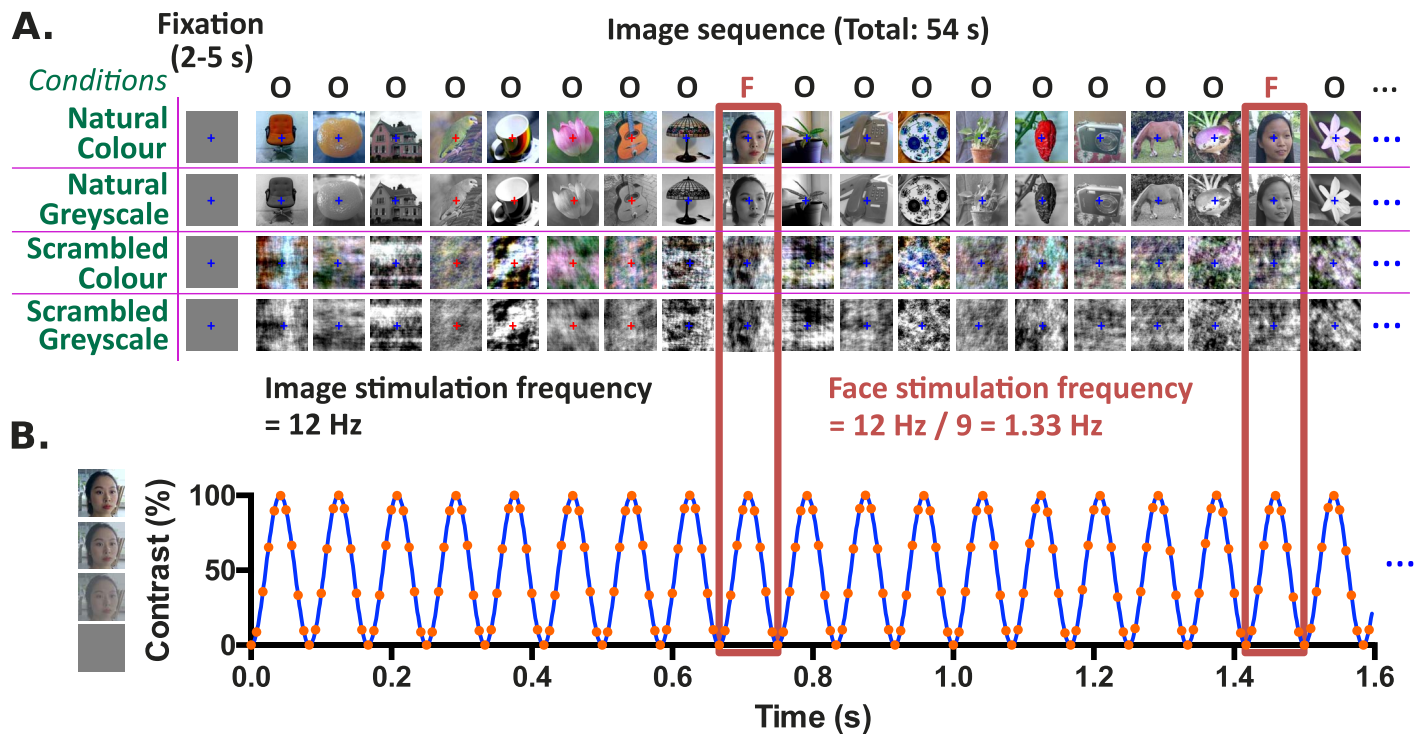


Figure 1. Procedure in Experiment 1. (A) In each condition, a stimulation sequence started with a brief fixation period followed by 648 images containing a random face (F) presented periodically after every presentation of eight nonface random objects (O) (i.e., one face every nine stimuli). In the scrambled image conditions, the fixed periodicity of face presentation remained but faces and nonface objects were replaced by their scrambled versions respectively. The participant's task was to press a key when the fixation cross changed color (blue to red for 300 ms; note that the color changes did not coincide with the onsets and offsets of images). Here, the figure shows the first 19 images identical across conditions for illustration purposes only. In actual experiments, each sequence contained a random array of images and random timings of fixation color change uncorrelated across conditions and observers, and included fade-in and fade-out periods (2 s each) not illustrated here (see text). (B) Each periodic stimulus (duration: 83.3 ms, i.e., 12.0 Hz frequency) was presented through a gradual increase and decrease of contrast over 10 frames (8.33 ms/frame at 120 Hz screen refresh rate; orange dot: onset time of a frame), following a sinusoidal contrast modulation (left: example stimuli at 0%, 36%, 65%, and 100% contrasts, bottom to top). The red boxes represent periodic presentations of face or scrambled face stimuli at 1.33 Hz. The face images shown here are for illustrations only and were not used in actual experiments.

direct, and robust signature of automatic natural-image face categorization (Rossion et al., 2015). By presenting faces at a fixed rate among nonface objects in rapid succession (Figure 1), a periodic electrophysiological response associated with the specific periodic face presentations necessarily reflects both direct discrimination of faces from many nonface objects (rather than measuring responses to different types of stimuli separately) and a generalized response across a wide range of face stimuli differing in lighting, viewpoint, face race, expressions, and so on. This paradigm has been validated in previous studies (e.g., De Heering & Rossion, 2015; Jacques, Retter, & Rossion, 2016; Jonas et al., 2016; Retter & Rossion, 2016; Rossion et al., 2015). Such a periodic response is best captured by characteristic, narrow peaks at the frequency of periodic face presentations and its harmonics in a spectral analysis of the EEG signals. Note that a significant periodic response emerges only from a response to

repeated face stimuli, in direct comparison with a differential response to other nonface stimuli in the sequence, while potentially confounding low-level visual cues are controlled by variability in the images (Rossion et al., 2015; see also Rossion, Jacques, & Jonas, 2018 for further review). Among these FPVS–EEG studies, part of them presented images in full color only, while the rest of them presented images in grayscale only. Thus, no prior FPVS–EEG study directly compared the face-selective responses to a full-color image sequence and to a grayscale image sequence in a single experimental design.

In order to directly examine the effect of color on face categorization, we presented the natural image sequences in two separate conditions (Figure 1), the first of which contained full color information across all images, and the other consisted of images all in grayscale. This would allow direct comparisons of the resulting FPVS–EEG data from the two conditions, in

terms of the amplitudes of the peaks in the frequency domain. Additionally, through a time-domain analysis of the EEG signal obtained during FPVS (Retter & Rossion, 2016), we explored whether the effect of color is early (i.e., more likely to affect segmentation and discrimination of faces) and/or later (i.e., more likely to affect later stages of face perception; see Gegenfurtner & Rieger, 2000; Yao & Einhäuser, 2008).

We also designed two extra conditions where phase-scrambled versions of the stimuli, in color in one condition and in grayscale in another condition, were rapidly displayed in the same settings as in the natural image conditions (i.e., one scrambled face every nine scrambled stimuli; Figure 1). Phase scrambling is a manipulation that preserves Fourier amplitude information carrying global low-level statistical properties of images, but removes the shape and structure in the stimuli (Sadr & Sinha, 2004). It has been typically used as a control for the contribution of low-level properties to object categorization, and face categorization in particular (e.g., Rossion et al., 2015; Torralba & Oliva, 2003; VanRullen, 2006), with studies showing that the earliest saccades toward natural images of faces in binary decision tasks can be significantly affected after controlling for the amplitude spectrum (Crouzet & Thorpe, 2011). Here, these additional conditions allowed investigation of potential low-level color contribution to face categorization.

Experiment 1

Methods

Participants

A total of 20 observers (10 females, mean age = 22.7 ± 3.2 years, age range: 19–36 years) participated in the experiment. All participants had normal or corrected-to-normal visual acuity. They were all right-handed according to an adapted Edinburgh Handedness Inventory measurement (Oldfield, 1971). None reported any history of psychiatric or neurological disorders. They were naive to the purpose of study, and were not aware that faces were presented at a fixed rate of one out of nine stimuli and that scrambled stimuli were generated from objects and faces. All participants provided written informed consent and received honoraria for their participation, as approved by the Biomedical Ethical Committee of the University of Louvain and the 2013 WMA Declaration of Helsinki.

Stimulus display

The stimuli were generated by a Dell XPS Desktop computer installed with the Psychtoolbox 3.0.8 in

MATLAB R2009a for Windows (MathWorks, Natick, MA) using previously validated scripts (e.g., Rossion & Boremanse, 2011), passed to a GeForce GTX 560 Ti graphics card, and were displayed on a linearly gamma-corrected BenQ XL2420T monitor at a refresh rate of 120 Hz, with a screen resolution of 1920 × 1080 pixels placed at a viewing distance of 80 cm (pixel size: 0.0194°) in a dimly lit and sound-attenuated room. The mean luminance after gamma correction was 75.0 cd/m².

Stimuli

Four types of images, detailed as follows, were generated for four corresponding conditions used in both experiments.

Natural color images: Color photographs of 46 faces and 247 nonface objects (animals, plants, man-made objects, houses, etc.; examples in Figure 1A) were obtained from the Internet (stimuli available here: <http://face-categorization-lab.webnode.com/resources/natural-face-stimuli/>). All faces (width = 1.9°–4.1°, height = 2.8°–5.0°) and objects, variable in size, lighting condition, and background, were located at the center of the square image stimuli (size: 5.1° × 5.1°) and embedded in their original natural scenes (i.e., unsegmented) after rescaling and cropping the source images. Each stimulus was coded and displayed in the RGB mode at a color depth of 24 bits/pixel (8 bits/channel, with pixel values of 0–255 representing luminance). The average luminance of each stimulus was equalized to the screen's mean luminance (75.0 cd/m²). Such luminance normalization was performed independently for each color channel, so that the average R, G, and B pixel values across the entire image were all normalized to 127.5 (mean pixel value). Note that variations in local color, luminance, and contrast within each image (i.e., appearances of the actual faces and objects) remained and the relative color and contrast variations across images were broadly maintained (Figure 1A). The luminance-normalized natural color images served as the basis for the generation of natural grayscale images and scrambled images.

Natural grayscale images: The luminance-normalized color images were each converted to grayscale using the formula:

$$I = 0.2126R + 0.7152G + 0.0722B, \quad (1)$$

where I is the pixel value representing luminance, and R , G , and B represent the original red, green, and blue values respectively. The weights in Equation 1 are standard for a gamma-corrected monitor (with color space following Rec. 709 primaries). Each stimulus was subsequently coded and displayed at a grayscale

resolution of 8 bits/pixel, of which the average luminance stayed unchanged at 75.0 cd/m².

Scrambled color images: To remove the shape information, we scrambled the images by randomizing the phase spectrum in the Fourier domain (Sadr & Sinha, 2004). A two-dimensional fast Fourier transform (FFT) converted each color image into a complex representation consisting of magnitude and phase components. The phase values were then replaced by those from an FFT of a randomly generated white-noise image of the same size. An inverse FFT was subsequently applied to the resulting map consisting of unchanged magnitudes and random phases in order to generate a phase-scrambled version of the image. The resulting scrambled image effectively became unrecognizable but kept the same frequency spectrum and average luminance value as the original natural image (Figure 1A).

Scrambled grayscale images: All scrambled color images were converted to 8-bit grayscale using Equation (1), maintaining the same average luminance values.

Procedure

Each experiment was a 2×2 factorial design studying the effects of color and scrambledness of the image stimulations. Thus, there were four conditions (Figure 1), each using a specific type of image described in Stimuli (i.e., natural color images, natural grayscale images, scrambled color images, and scrambled grayscale images). In each condition, the stimulation sequence was presented through sinusoidal contrast modulation (Figure 1B; e.g., Jacques et al., 2016; Retter & Rossion, 2016; Rossion et al., 2015) of successive images at a rate of 12.0 Hz (image stimulation frequency). Each 83.3-ms (1000 ms / 12.0, 10 frames/image) stimulation cycle started with a uniform gray background from which an image appeared as its contrast increased in a sinusoidal fashion from 0%, reaching 100% (full contrast) at 41.7 ms, and then decreased at the same rate. In the natural image conditions, the periodic sequence comprised eight objects (O) followed by a face (F), all randomly selected from their corresponding categories. Similarly, in the scrambled image conditions, the periodic sequence consisted of eight scrambled objects followed by a scrambled face. Faces (or, scrambled faces) were thus presented at a frequency of 12.0 Hz / 9 = 1.33 Hz (face stimulation frequency). Images could be repeated one to three times randomly (but not consecutively) within a stimulation sequence.

A stimulation sequence started with a fixation cross (in blue, $0.31^\circ \times 0.31^\circ$) centered on a uniform gray background for 2–5 s (duration randomly determined

across sequences) to facilitate stable fixation of the participant. The stimulation sequence, consisting of 648 images, was subsequently presented centrally on the screen for 54.0 s, including a 2-s fade-in period at the beginning of image presentation and a 2-s fade-out period at the end (with uninterrupted central display of the fixation cross superimposed on the images). The contrast modulation depth of the periodic stimulation gradually increased from 0% to 100% during the fade-in period, and reduced in the opposite direction from 100% to 0% during the fade-out period (keeping the sinusoidal contrast modulation). These fading periods were intended to minimize blinks and abrupt eye movements due to an otherwise sudden appearance or disappearance of the flickering stimuli. Responses during the fading periods would not be used in the data analyses, as detailed later.

Each participant performed 12 sequences (three per condition), each of which contained an independently randomized image sequence. The order of presentation of the 12 presentation sequences was randomized. During the EEG recording, the participant was instructed to maintain central fixation throughout the entire stimulation sequence while continuously monitoring the flickering stimuli. As in previous studies with this paradigm (e.g., Rossion et al., 2015), the participants' task was to detect brief color changes of the fixation cross (blue cross to red cross for 300 ms; i.e., 36 frames). Such color changes occurred 10 times randomly throughout each sequence, and were not correlated with the onsets and offsets of images. The accuracy (hit rate: percentage that the observer correctly pressed the key within 1500 ms after the onset of the color change) and response times for accurate key presses were analyzed.

EEG acquisition

The EEG was acquired using a 128-channel Biosemi Active 2 system (BioSemi, Amsterdam, The Netherlands), with electrodes including standard 10–20 system locations as well as additional intermediate positions (<http://www.biosemi.com/headcap.htm>, re-labeled to more conventional labels of the 10–5 system; see supplementary figure S1 in Rossion et al., 2015). The EEG was sampled at 512 Hz. Electrode offset was reduced to under ± 20 mV for each individual electrode by softly abrading the scalp underneath with a blunt plastic needle and injecting the electrode with saline gel. Eye movements were monitored by four additional electrodes placed at the outer canthi of the two eyes, and above and below the right orbit. During the experiment, triggers were sent via parallel port from the stimulation computer to the EEG recording computer at the beginning and the end

of each stimulation sequence, and at the minima (0% contrast) of all 12.0-Hz stimulation cycles (i.e., onsets/offsets of images), using custom scripts borrowing from the Cogent 2000 MATLAB Toolbox (validated in previous studies, e.g., Rossion & Boremanse, 2011). The temporal synchrony between the trigger and the stimulus onset was verified by a photodiode prior to the experiment. Recordings were manually initiated by the experimenter when participants showed artefact-free EEG signals.

EEG analysis

Preprocessing

All EEG data were analyzed using Letswave 5 (<http://nocions.webnode.com/letswave>) running on MATLAB. The signals were first detrended by subtracting the best-fit line (using the least-squares method) from the data, and then passed to a fourth-order low-pass Butterworth filter (Butterworth, 1930) with a cutoff frequency of 120 Hz. The data were then passed to an FFT multinotch filter (width = 0.5 Hz) to remove electrical noise at 50 Hz (oscillation frequency of the alternating current) and its second harmonic (100 Hz). Subsequently, the filtered signals were segmented into 58-s segments, keeping 2 s each before and after a sequence (i.e., –2 s through 56 s). The DC component in each data segment was separately identified and then subtracted from the signal.

Artefacts in the signals were removed in two steps. Blink artefacts were removed only when a participant's blink rate exceeded 0.2 blink/s (Retter & Rossion, 2016), resulting in only one participant meeting this criterion. An independent component analysis (Jung et al., 2000) using the square mixing matrix method was subsequently applied on this participant's signals and only one single component connected to the blink patterns was removed, chosen based on visual inspection of the waveform and its topography. Then, noisy and artefact-ridden channels (fewer than 5% of 128 channels; i.e., a maximum of six channels) containing deflections larger than 100 μ V in multiple presentation sequences were rebuilt using linear interpolations from immediately adjacent noise-free channels. Finally, all channels (except the ocular ones) were referenced to a common average.

Frequency-domain analysis

The preprocessed data segment of each sequence was cropped again to keep only signals from exactly 2 s after stimulus onset (the end of the fade-in period) to 51.5 s after stimulus onset. The end time (51.5 s) was chosen such that it was the longest possible time point before the start of stimulus fade-out (at 52 s), for

capturing an integer number of 1.33-Hz cycles (i.e., $1.33 \text{ Hz} \times 49.5 \text{ s} = 66$ cycles, which contains $N = 25,348$ time bins). The integer number of cycles ensured no spectral leakage of the frequencies of interest—that is, harmonics of both the face stimulation frequency (1.33 Hz) and the image stimulation frequency (12.0 Hz). The sequences were then averaged separately for each condition and for each observer. An FFT was applied to the sequence-averaged data segments, and an amplitude spectrum (normalized by $N/2$, in μ V) was extracted in the frequency domain (ranging from 0 to 256 Hz) for each channel. Each spectrum had a high frequency resolution (i.e., distance between two adjacent frequency components) of 0.0202 Hz, which is the inverse of the segment duration (49.5 s). This aided unambiguous identification of the frequencies of interest (1.33 Hz and harmonics).

To consider the variations of noise across the amplitude spectrum, a baseline subtraction was applied to each frequency component by subtracting the average amplitude of 20 surrounding frequency components (10 on each side, excluding the immediately adjacent bins and the local minimum and maximum bins; see, for example, Dzhelyova & Rossion, 2014; Mouraux et al., 2011) from the amplitude of the frequency component of interest. In addition, the signal-to-noise ratio (SNR) was also calculated by considering the same 20 surrounding frequency components (e.g., Rossion, Alonso Pireto, Boremanse, Kuefner, & Van Belle, 2012). For group analysis, individual baseline-subtracted amplitude (or SNR) spectra were averaged across observers for each condition, resulting in the grand-averaged spectrum.

Selecting the range of significant harmonic responses (z-score analyses): To analyze the responses at the image stimulation and face stimulation frequencies (and their harmonics), we first determined a continuous range of significant harmonic responses for each frequency to include in the analysis. Individual amplitude spectra were first averaged across observers, and then across the 128 channels (excluding the four ocular channels) for each condition. A z-score was calculated for each frequency component of this averaged spectrum by using the mean amplitude and standard deviation of 20 surrounding frequency components (10 on each side, excluding the immediately adjacent bins; see Rossion et al., 2012) from the amplitude of the frequency component of interest. For face stimulation responses, the harmonics to be included in the analysis ranged from 1.33 Hz through a cutoff frequency determined by the last significant harmonic that yielded a z-score larger than 2.33 (i.e., beyond the 99.0 percentile of the SNR distribution; Retter & Rossion, 2016) in the two natural image conditions, as no significant face stimulation responses were expected for scrambled image conditions (Rossion et al., 2015). Similarly for

image stimulation responses, the included harmonics (from 12.0 Hz) were determined by the same *z*-score criterion but considering all four conditions. The significant harmonic responses were summed, separately for each frequency type, in order to quantify and compare the comprehensive response amplitudes and scalp topographies across conditions (Retter & Rossion, 2016).

Statistical comparisons across conditions: A 2 (Natural vs. Scrambled) \times 2 (Color vs. Grayscale) repeated-measures analysis of variance (ANOVA) was performed on baseline-subtracted amplitudes (summed over significant harmonics) averaged over all 128 channels for each of the 20 observers. We also defined regions of interest (ROIs) over occipitotemporal and occipitoparietal channels that showed the largest responses, and analyzed these responses in additional ANOVAs in order to localize the potential color advantage in the brain.

Time-domain analysis

The periodic responses were additionally examined in the time domain (e.g., Dzhelyova & Rossion, 2014; Jacques et al., 2016; Retter & Rossion, 2016; Rossion et al., 2015). The preprocessed data segments were each passed to a fourth-order bandpass Butterworth filter with a bandwidth of 0.1–30 Hz. The choice of cutoff at 30 Hz was based on previous studies (e.g., Jacques et al., 2016; Retter & Rossion, 2016) and standard procedures for ERP analyses that investigated face-selective responses (e.g., Rossion & Caharel, 2011; Rousselet, Husk, Bennett, & Sekuler, 2007; see the review of Rossion & Jacques, 2008), which covered and went beyond the entire range of significant harmonics (up to 16.0 Hz) of face categorization responses in the current study (see EEG data: Frequency-domain analysis). The filtered data segment was further cropped to keep only signals from stimulus onset (0 s) to 51.9 s after. The end time (51.9 s) was chosen such that it was the nearest time point to the start of stimulus fade-out (at 52 s) for capturing an integer number of 12.0-Hz cycles (i.e., 12.0 Hz \times 51.9 s = 623 cycles, which contains $N = 26,586$ time bins). An FFT multinotch filter (width = 0.5 Hz) was subsequently applied to the cropped signals to selectively remove 12.0 Hz and its first three harmonics, corresponding to the contribution of the base stimulation to the time-domain waveforms. The filtered signals were then cropped into smaller epochs of 1417 ms (17 \times 83.3-ms base stimulation cycles), each including responses to a sequence of eight object stimuli, one face stimulus, and another eight object stimuli (OOOOOOOOF00000000). Thus, each epoch contains responses for exactly one face stimulus. The cropping began at 2.25 s after stimulus onset,

which was the earliest time point possible after the 2-s fade-in period. It should be noted that the first eight object stimuli of each epoch correspond to the last eight object stimuli of its immediately preceding epoch. After averaging all epochs per observer for each condition, the data were baseline-corrected by subtracting the mean response amplitude across 167 ms (corresponding to the presentation of two object stimuli) preceding presentation of the face stimulus in the epoch sequence. For each condition, the baseline-corrected responses for all 20 participants were subjected to a two-tailed *t* test at each time point. A face-selective component was defined by a time window where significant nonzero responses ($p < 0.05$) were found across 12 or more consecutive time points (i.e., ≥ 21.5 ms; see, e.g., Jacques et al., 2016; Laganaro, 2014). Similar statistical treatment was applied to the within-subjects difference of individual baseline-corrected responses between natural color image and natural grayscale image conditions in order to evaluate any potential color advantage in face categorization in the time domain.

Results

Behavioral data

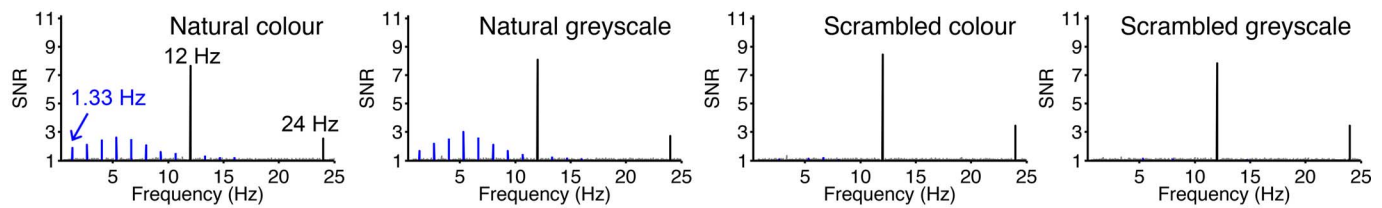
Observers' accuracy (hit rates) and response times for accurate key presses over 30 color changes (10 changes \times 3 sequences per condition) were analyzed. The mean hit rates were close to ceiling in all conditions (all over 95% correct: natural color: 95.6% \pm 1.6%, natural grayscale: 97.6% \pm 1.1%, scrambled color: 98.4% \pm 0.6%, scrambled grayscale: 97.5% \pm 0.7%; all in $M \pm 1$ SEM). The response times were rapid (< 500 ms for all conditions) but varied slightly across conditions (natural color: 464 ms \pm 11 ms, natural grayscale: 445 ms \pm 11 ms, scrambled color: 441 ms \pm 11 ms, scrambled grayscale: 433 ms \pm 11 ms; all in $M \pm 1$ SEM). A 2 \times 2 repeated-measures ANOVA on response times showed significant differences for both main effects (color $>$ grayscale, 3% effect: $F(1, 19) = 6.86$, $p = 0.02$; natural $>$ scrambled, 4% effect: $F(1, 19) = 22.4$, $p < 0.001$) but no significant interaction effect, $F(1, 19) = 3.12$, $p = 0.09$.

EEG data

Frequency-domain analysis

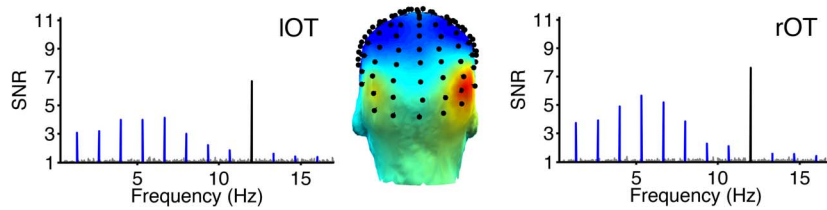
Here, we report responses to the frequency rates that represent, respectively, face stimulation (1.33 Hz and harmonics) and image stimulation (12.0 Hz and harmonics).

A. Frequency spectra (Averaged across all observers & all channels)

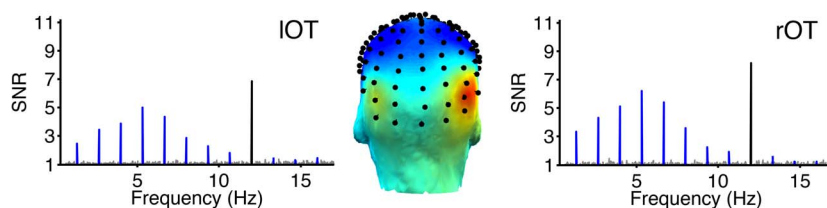


B. Face stimulation responses (1.33 Hz & harmonics)

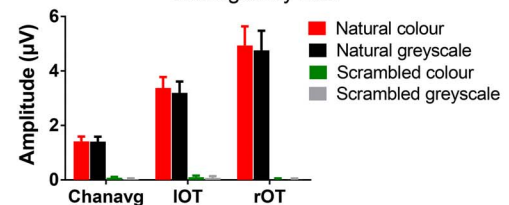
Natural colour



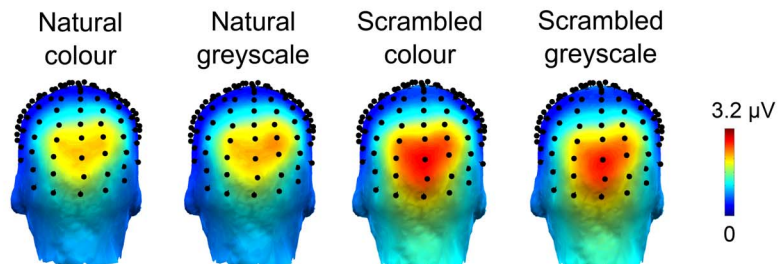
Natural greyscale



Harmonic-summed responses averaged by ROI



C. Image stimulation topographies (12 Hz & harmonics)



D. ROI definitions

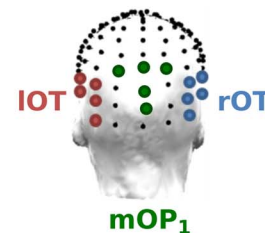


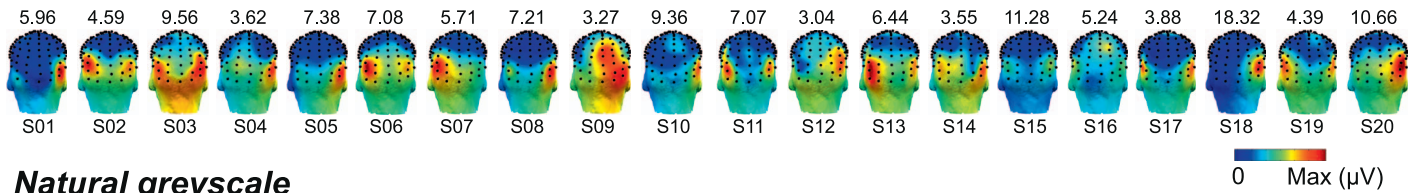
Figure 2. Experiment 1: Frequency-domain responses. (A) For each of the four conditions, the frequency spectrum plots the SNR averaged over all observers and all 128 channels as a function of frequency. Black lines: image stimulation responses (12 Hz and harmonics). Blue lines: face stimulation responses (1.33 Hz and harmonics). (B) Responses to face stimulation, focusing on the two natural image conditions. Each frequency spectrum shows the SNR averaged over observers and IOT/rOT channels as a function of frequency. For each condition, the scalp topography (back of the head) shows the sum of observer-averaged, baseline-subtracted amplitudes over significant face stimulation harmonics (1.33–16 Hz, except 12 Hz). The bar graph shows the harmonic sums of baseline-subtracted amplitudes averaged over all 128 channels (Chanavg), IOT and rOT channels separately for all conditions. Each bar represents the mean over 20 observers (error bar = 1 SEM). (C) Responses to image stimulation. Each scalp topography shows the sum of observer-averaged, baseline-subtracted amplitudes across significant image stimulation harmonics (12–60 Hz). (D) Corresponding channel locations that define the IOT, rOT, and mOP₁ ROIs.

Face stimulation frequency (1.33 Hz)

Average across all channels: Figure 2A shows the frequency spectra (in form of SNRs) for mean responses over all 128 channels and all 20 observers. Robust responses were observed at face stimulation frequency (1.33 Hz) and its harmonics only for the

natural image conditions (blue lines), representing the brain's discrimination of faces from other objects (i.e., face-selective responses) only when the shape information was intact. A *z*-score analysis (see Methods) was performed on the averaged spectra for the natural image conditions in order to determine the range of

Natural colour



Natural greyscale

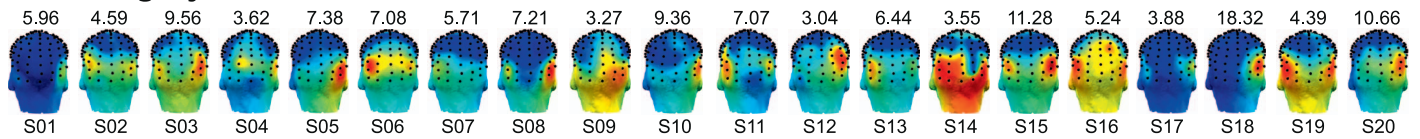


Figure 3. Experiment 1: Individual frequency-domain scalp topographies for the two natural image conditions. A back-of-the-head topography shows the sums of baseline-subtracted amplitudes across significant face stimulation harmonics (1.33–16.0 Hz, except 12 Hz) for each of the 20 participants. The color scale is identical across conditions within each participant, but the maximum amplitude (on top of each topography) varies across participants.

relevant harmonics. The highest significant harmonics (i.e., z score > 2.33) were both 16.0 Hz (12th harmonic). (Note: Over the same range, only one harmonic was significant in each of the scrambled image conditions.) The baseline-subtracted amplitudes for each observer and condition were subsequently summed across these significant harmonics (i.e., over the range of 1.33, 2.67, 4.00 Hz, and so on until 16.0 Hz, excluding 12.0 Hz, which coincides with the image stimulation frequency; see grand-averaged scalp topographies in Figure 2B) for the following analyses.

To compare the responses across conditions, the individual harmonic-summed, baseline-subtracted amplitudes, further averaged over all 128 channels (Bar graph in Figure 2B: Chanavg), were subjected to a 2×2 repeated-measures ANOVA. A significant main effect was found for natural versus scrambled conditions, $F(1, 19) = 76.0$, $p < 0.001$. Precisely, responses to scrambled images were only 4.03% of those to natural images on average. Importantly, however, no significant main effect was found for color versus grayscale conditions, $F(1, 19) = 0.23$, $p = 0.64$, nor a significant interaction, $F(1, 19) = 0.18$, $p = 0.67$. Thus, when considering a data average over all channels, we did not find that color's presence enhanced the face categorization response. *Occipitotemporal regions:* To understand the spatial distribution of the face stimulation responses across the scalp, the harmonic-summed, baseline-subtracted amplitudes were averaged across observers for each condition. The observer-averaged scalp topographies (Figure 2B) revealed the largest responses to natural faces (in both color and grayscale) over the occipitotemporal regions, lateralized to the right hemisphere. As in previous studies with this approach, significant face-selective responses were found in every single participant and every condition with natural shapes, and individual scalp topographies (Figure 3) also suggest mainly occipitotemporal responses. To define the ROIs, we ranked the channels according to their mean responses

over the two natural image conditions. We then defined the right occipitotemporal ROI (rOT) by the top five channels over this area (i.e., P10, PO10, PO12, PO8, and P8; Figure 2D). (These top channels were identical when considering the two conditions separately.) The left occipitotemporal ROI (lOT) was defined as the symmetrical channels in the left hemisphere (i.e., P9, PO9, PO11, PO7, and P7; Figure 2D). To compare the responses across conditions (bar graph in Figure 2B), a 2 (Color vs. Grayscale) $\times 2$ (Natural vs. Scrambled) $\times 2$ (lOT vs. rOT) repeated-measures ANOVA was performed on the individual harmonic-summed, baseline-subtracted amplitudes averaged over corresponding ROI channels. Similar to results for 128-channel averages, the main effect of natural versus scrambled conditions (NS) was significant, $F(1, 19) = 93.0$, $p < 0.001$, where the average responses to scrambled images were only 1.13% of those to natural images. However, the main effect of color versus grayscale conditions (CG) was not significant, $F(1, 19) = 0.42$, $p = 0.52$. The main effect of ROI was almost significant, $F(1, 19) = 4.22$, $p = 0.054$. No interaction terms were significant, CG \times NS: $F(1, 19) = 0.37$, $p = 0.55$, CG \times ROI: $F(1, 19) = 0.05$, $p = 0.83$, CG \times NS \times ROI: $F(1, 19) = 0.03$, $p = 0.86$, except for the significant interaction of NS \times ROI, $F(1, 19) = 5.14$, $p = 0.04$. We subsequently conducted post hoc pairwise comparisons (a) between responses to natural and scrambled images, and (b) between the two ROIs. Not surprisingly, all pairwise comparisons showed significantly larger responses to natural images than scrambled images. Interestingly, responses were significantly right lateralized only for the natural image conditions (rOT $>$ lOT by 46% and 49% for the color and grayscale conditions, $p = 0.047$ and $p = 0.041$, respectively) but not for the scrambled image conditions ($p = 0.11$ and $p = 0.38$ for the color and grayscale conditions, respectively). Overall, the results suggested a high-level, right-lateralized occipitotemporal response for face categorization,

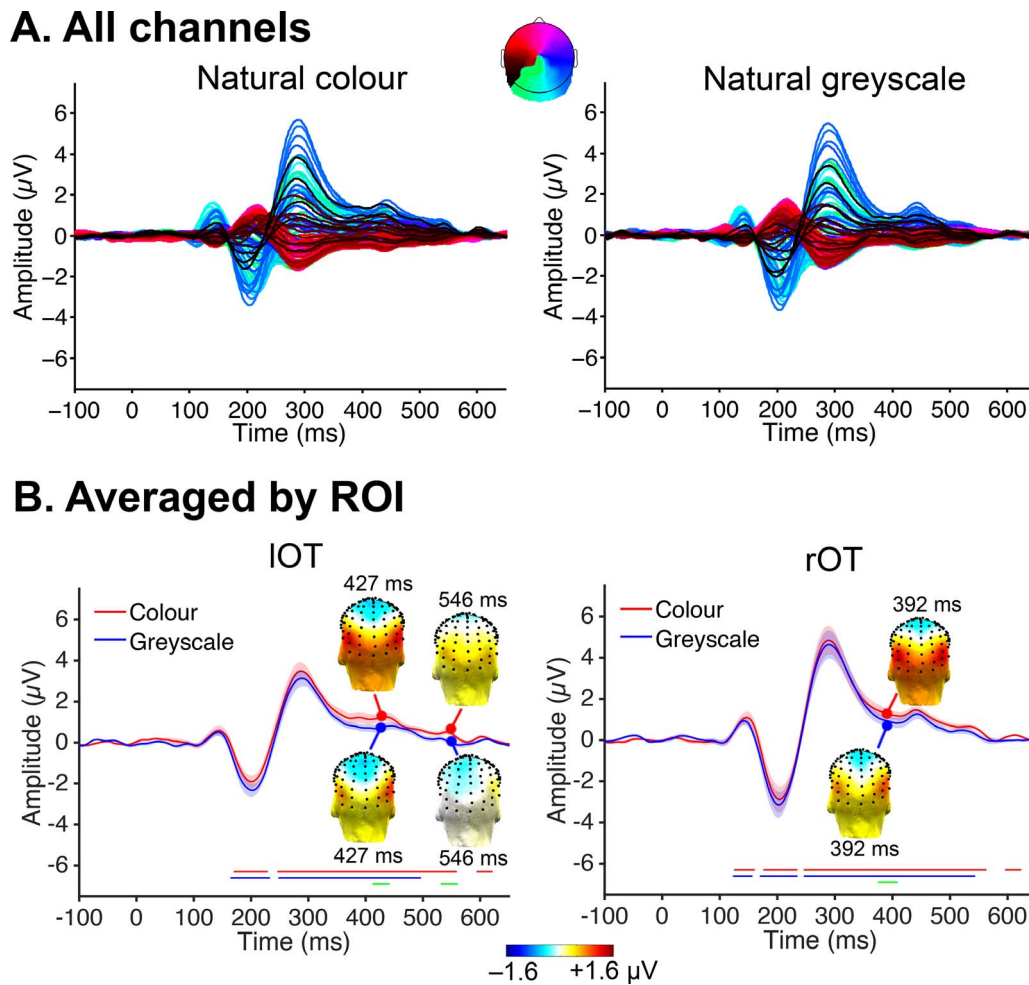


Figure 4. Experiment 1: Time-domain responses to face stimulation in the two natural image conditions. Periodic data were segmented relative to the onset of face stimulation (0 s), notch-filtered at 12 Hz and harmonics, averaged across data segments and observers, and baseline-corrected in order to reveal waveforms associated with face stimulation (see text). (A) Waveforms for all 128 channels. The two-dimensional head map (viewed from top of the head) represents the color codes for the channels. (B) Waveforms averaged by ROI (see definitions in Figure 2D). Shaded areas represent ± 1 SEM across observers. For each ROI, the bottom horizontal lines represent significantly nonzero responses over 12 consecutive time points (i.e., $p < 0.05$ for 21.5 ms), respectively, to natural color images (red), natural grayscale images (blue), and the difference between the two conditions (green). At the midpoint of each green bar, the scalp topographies (back of the head) reveal superior face stimulation responses when the images contained color. The color scales are identical across all scalp topographies.

but failed to show significantly larger responses when the natural images contained color.

Image stimulation frequency (12.0 Hz)

Average across all channels: A response at the image stimulation frequency (12.0 Hz) and its harmonics was found for all four conditions (black bars in Figure 2A). This response merely represents the brain's sensitivity to the rate of image stimulations regardless of object image type and is not the focus of the study. Nevertheless, to evaluate the image stimulation responses across conditions, we conducted a z -score analysis on the averaged spectra (see Methods) in order

to determine the range of relevant harmonics to be considered. The analysis revealed that the first five harmonics (12.0, 24.0, 36.0, 48.0, and 60.0) were significant (i.e., z score > 2.33) in the natural grayscale image condition, while the first four harmonics were significant in all other conditions. Thus, for each observer and condition, we decided to compute the sum of baseline-subtracted amplitudes across the first five harmonics (see scalp topographies in Figure 2C) in the following analyses. It should be noted that amplitudes at nonsignificant harmonics were close to zero, and adding these values would not change the result of the study.

In order to compare the responses across conditions, the individual harmonic-summed, baseline-subtracted

amplitudes were first averaged over all 128 channels, and then subjected to a 2×2 repeated-measures ANOVA. A significant main effect was found for natural versus scrambled conditions, $F(1, 19) = 5.14$, $p = 0.04$, where responses to scrambled images were on average 13.8% larger than responses to natural images. However, we found no significant main effect for color versus grayscale conditions, $F(1, 19) = 0.034$, $p = 0.86$, and also no significant interaction, $F(1, 19) = 2.12$, $p = 0.16$.

Medial occipitoparietal area: To spatially localize the responses, the harmonic-summed, baseline-subtracted amplitudes were averaged across observers for each condition (Figure 2C). An obvious observation is that responses peaked over the medial occipitoparietal area consistently across the four conditions, though with apparently varying magnitudes. To evaluate such responses, we defined the medial occipitoparietal ROI (mOP₁) by first ranking the channels according to their responses averaged over the four conditions, and then selecting the five channels that scored the highest (i.e., Oz, POO6, Oiz, POO5, and POOz; Figure 2D). As an additional note, these five channels were among the top eight consistently in the four separate conditions. The individual harmonic-summed, baseline-subtracted amplitudes averaged over the five channels were subjected to a 2×2 repeated-measures ANOVA. Similar to the results for averaging over all channels, responses to scrambled images were significantly larger than those to natural images by 16.1% (natural versus scrambled conditions: $F(1, 19) = 6.93$, $p = 0.02$). No significant main effect for color versus grayscale conditions was observed, $F(1, 19) = 0.26$, $p = 0.62$, nor was there a significant interaction, $F(1, 19) = 3.76$, $p = 0.07$. These results suggest that variations in image stimulation responses were largely driven by activities over mOP₁.

Time-domain analysis

Figure 4A shows the time-domain responses, in terms of baseline-corrected amplitudes averaged across all epochs and observers for all 128 channels in the two natural image conditions after selectively notch-filtering out the image-stimulation rate response (see Methods; further details in Retter & Rossion, 2016). Differential waveforms were time-locked to the periodic face stimuli (onset: 0 s) reflecting a face-selective process, regardless of the presence of color in the images. Qualitatively, we observed at least three distinctive components underlying the face-selective responses over time: P1-face, N1-face, and P2/P3-face. The timings of these components were generally consistent across the two natural image conditions, and agreed with previous findings (e.g., Retter & Rossion, 2016). For responses averaged over rOT channels (Figure 4B), in particular, we defined at least three

components (red and blue horizontal lines near the bottom of the plot) by significant nonzero responses ($p < 0.05$) over 12 consecutive time points (i.e., 21.5 ms). In the natural color image condition, we observed a small P1-face component at 126–159 ms after stimulus onset, then an N1-face component at 177–234 ms, followed by a large P2/P3-face component starting from 247 ms that remained positive until 562 ms (and finally, a small positive component at 597–622 ms). The timings of components in the natural grayscale image condition were similar—P1-face: 124–156 ms, N1-face: 171–234 ms, P2/P3-face: 247–542 ms. Responses over the IOT were similar, except that the P1-face was not significant—Color: N1-face: 171–228 ms, P2/P3-face: 247–558 ms (and a small positive component at 595–620 ms similar to that over the rOT), grayscale: N1-face: 165–232 ms, P2/P3-face: 249–495 ms.

To examine any potential effect of adding color in the natural images, we subtracted individual ROI-averaged response waveforms between the natural grayscale image and the natural color image conditions (significant differences over 12 consecutive time points, or 21.5 ms, in green lines in Figure 4B, $p < 0.05$, two-tailed t test). A significantly larger response to color than to grayscale images was found at relatively late latencies, far beyond the onset of P2/P3-face component. For the rOT, a significant color advantage was found only between 376 and 407 ms after stimulus onset. For the IOT, a significant color advantage occurred even later but in two small, separate time intervals: 413–441 ms and 532–560 ms. Thus, the time-domain analysis revealed a small, late, but significant advantage in the presence of color.

Discussion

In this experiment, as in previous studies with this paradigm (e.g., Retter & Rossion, 2016; Rossion et al., 2015), we obtained robust (i.e., significant for all participants) face categorization EEG responses to natural images, with the largest response found over the right occipitotemporal cortex. Strikingly, we found no significant global advantage in the frequency domain for color images over grayscale images. However, the time-domain analysis revealed a small, late, and sustained color advantage. A potential caveat is that the presence of color in the stimuli slightly (i.e., 3%), but significantly slowed down observers' responses to a task that is intended to be orthogonal to face categorization: detecting the color changes in the fixation cross. One possibility is that the task itself, which involved a response regarding color, might have interacted with the perception of color in the stimuli (see also Zhu et al., 2008). This might have subsequently reduced (and delayed) a potential color

advantage in face categorization. To test this, we designed an additional experiment by replacing the previous task with detection of shape changes of the fixation cross.

Experiment 2

Methods

A new group of 20 observers (10 females) who were not involved in Experiment 1 participated in this experiment. They were chosen based on the same criteria as before. All experimental conditions were identical to Experiment 1, except that observers were instructed to press a key when the blue fixation cross briefly changed shape to a square outline (without any color change). Blink artefacts were removed only in one participant's data (blink rate > 0.2 blink/s) using the same procedure as in Experiment 1.

Results

Behavioral data

As in Experiment 1, observers' accuracy (hit rates) and response times for accurate key presses over 30 shape changes (10 changes \times 3 sequences per condition) were analyzed. The average hit rates were again higher than 95% in all conditions, indicating a ceiling effect (natural color: 95.3% \pm 1.4%, natural grayscale: 96.5% \pm 0.9%, scrambled color: 97.1% \pm 1.0%, scrambled grayscale: 97.7% \pm 0.7%). The response times were also less than 500 ms for all conditions (natural color: 476 ms \pm 8 ms, natural grayscale: 480 ms \pm 10 ms, scrambled color: 439 ms \pm 9 ms, scrambled grayscale: 431 ms \pm 8 ms). A 2×2 repeated-measures ANOVA showed no significant differences in response time between the color and grayscale conditions, $F(1, 19) = 0.22$, $p = 0.65$, but significantly shorter response times towards scrambled images than natural images, 10% effect, $F(1, 19) = 44.8$, $p < 0.001$. The interaction effect was not significant, $F(1, 19) = 0.78$, $p = 0.39$. Compared to the interference of detecting fixation color changes in Experiment 1, the presence of color in the images did not significantly affect the detection of the shape changes of the fixation cross. Thus, the shape task was orthogonal to the investigation of a potential color advantage in face categorization and was a more appropriate task for the research question. Interestingly, the shape contained in the natural stimuli slowed down the detection of fixation shape changes (10% effect for natural vs. scrambled) more than the detection of fixation color changes (4% effect in Experiment 1), possibly due to a stronger

interaction between the task and the stimuli both involving shape perception.

EEG data

Frequency-domain analysis

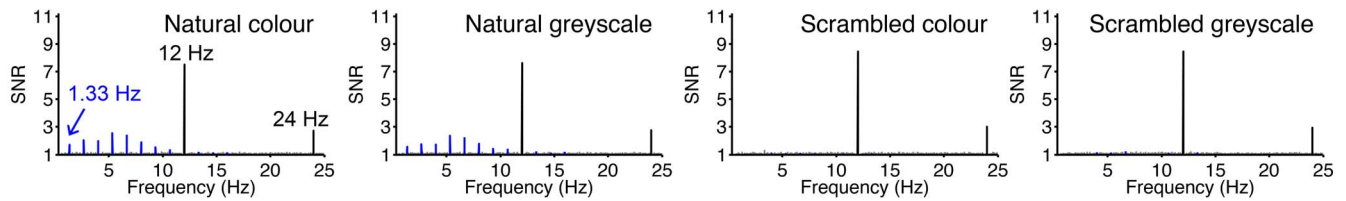
Face stimulation frequency (1.33 Hz)

Average across all channels: Figure 5A (blue lines) shows robust responses at face stimulation harmonics only for the natural image conditions. A z-score analysis revealed that the first 10 harmonics were significant for the natural grayscale image condition, and the first 12 harmonics were significant for the natural color image condition. Thus, the baseline-subtracted amplitudes were summed over the range of the first 12 harmonics (i.e., 1.33–16.0 Hz, except 12.0 Hz), consistent with the corresponding harmonic range in Experiment 1.

A 2×2 repeated-measures ANOVA on channel-averaged, harmonic-summed, baseline-subtracted amplitudes (bar graph in Figure 5B: Chanavg) showed significance for both main effects and the interaction effect, color vs. grayscale: $F(1, 19) = 5.68$, $p = 0.03$; natural vs. scrambled: $F(1, 19) = 68.6$, $p < 0.001$; interaction: $F(1, 19) = 5.11$, $p = 0.04$. Post hoc pairwise comparisons revealed that the natural color image condition scored a significantly larger response by 21.6% ($p < 0.02$) than the natural grayscale image condition, but no significant differences were found between the two scrambled image conditions ($p = 0.79$). On the other hand, as expected, the natural image conditions scored significantly larger responses than their corresponding scrambled image conditions ($p < 0.001$ in both comparisons; average response to scrambled images 6.1% of that to natural images). Overall, the results here suggested that, different from frequency-domain results in Experiment 1, the presence of image color showed a significant advantage for face categorization.

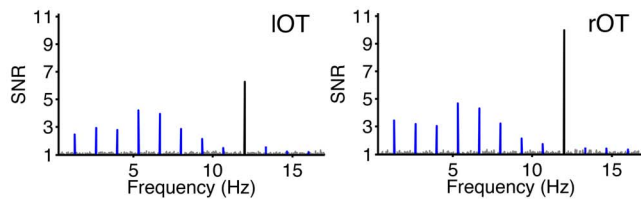
Occipitotemporal regions: As in Experiment 1, the observer-averaged topographies (Figure 5B) also revealed peak face categorization responses over the occipitotemporal areas, as evident for most individual observers (Figure 6). Ranking responses (averaged over the two natural image conditions) by channel showed that PO10, P10, PO12, P8, and PO8 scored the largest responses, and together defined the rOT ROI (Figure 5D), encompassing exactly the same channels as in Experiment 1. The ranking of the channels were indeed highly consistent between the two conditions, with the same four channels (i.e., P10, PO10, PO12, and PO8) at the top in both cases (P8 ranked 7th in the color condition, just behind PO9 and PO11; in the grayscale condition, P8 ranked 13th). The symmetric lOT was, as in Experiment 1, defined to encompass P9, PO9, PO11, PO7, and P7 (Figure 5D). A 2 (Color vs. Grayscale) \times 2

A. Frequency spectra (Averaged across all observers & all channels)

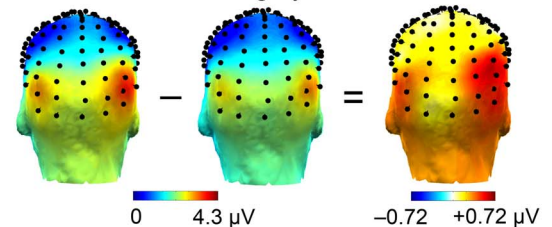


B. Face stimulation responses (1.33 Hz & harmonics)

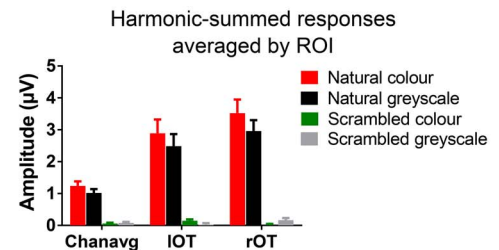
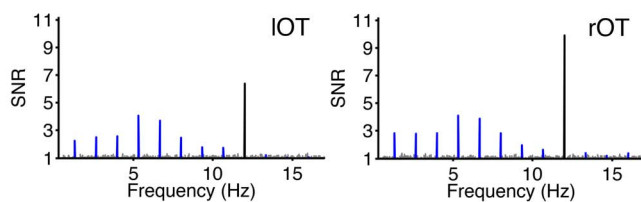
Natural colour



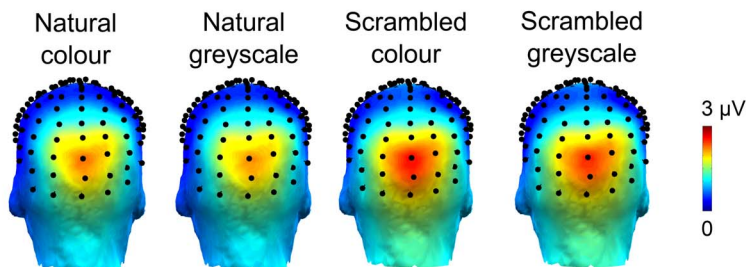
Natural colour Natural greyscale Difference



Natural greyscale



C. Image stimulation topographies (12 Hz & harmonics)



D. ROI definitions

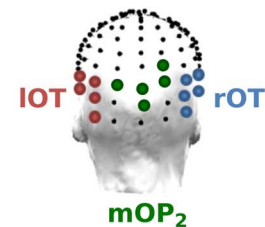


Figure 5. Experiment 2: Frequency-domain responses. (A) Each frequency spectrum plots the SNR averaged over all observers and all 128 channels as a function of frequency. Black lines: image stimulation responses (12 Hz and harmonics). Blue lines: face stimulation responses (1.33 Hz and harmonics). (B) Responses to face stimulation in the two natural image conditions. Each frequency spectrum shows the SNR averaged over observers and IOT/rOT channels as a function of frequency. The scalp topographies (back of the head) show the sums of observer-averaged, baseline-subtracted amplitudes over significant face stimulation harmonics (1.33–16.0 Hz) respectively for the natural image conditions (left two topographies), and their difference (rightmost topography). The bar graph shows the harmonic sums of baseline-subtracted amplitudes averaged over all 128 channels (Chanavg), IOT and rOT channels respectively, showing a significant colour advantage. Each bar represents the mean over 20 observers (error bar = 1 SEM). (C) Responses to image stimulation. Each scalp topography shows the sum of observer-averaged, baseline-subtracted amplitudes across image stimulation harmonics (12–60 Hz). (D) Corresponding channel locations that define the IOT, rOT, and mOP₂ ROIs. Note the difference of mOP₂ from mOP₁ (Figure 2D).

(Natural vs. Scrambled) \times 2 (IOT vs. rOT) repeated-measures ANOVA revealed that all three main effects were significant, color > grayscale: $F(1, 19) = 10.5, p = 0.005$; natural > scrambled: $F(1, 19) = 58.4, p < 0.001$; rOT > IOT: $F(1, 19) = 7.69, p = 0.013$. All interaction terms were significant, CG \times NS: $F(1, 19) = 6.45, p =$

0.02; NS \times ROI, $F(1, 19) = 8.35, p = 0.01$; CG \times NS \times ROI: $F(1, 19) = 7.38, p = 0.014$, except for the nonsignificant interaction of CG \times ROI: $F(1, 19) = 0.34, p = 0.57$. We subsequently conducted post hoc pairwise comparisons for all three parameters. Importantly, natural color images resulted in significantly

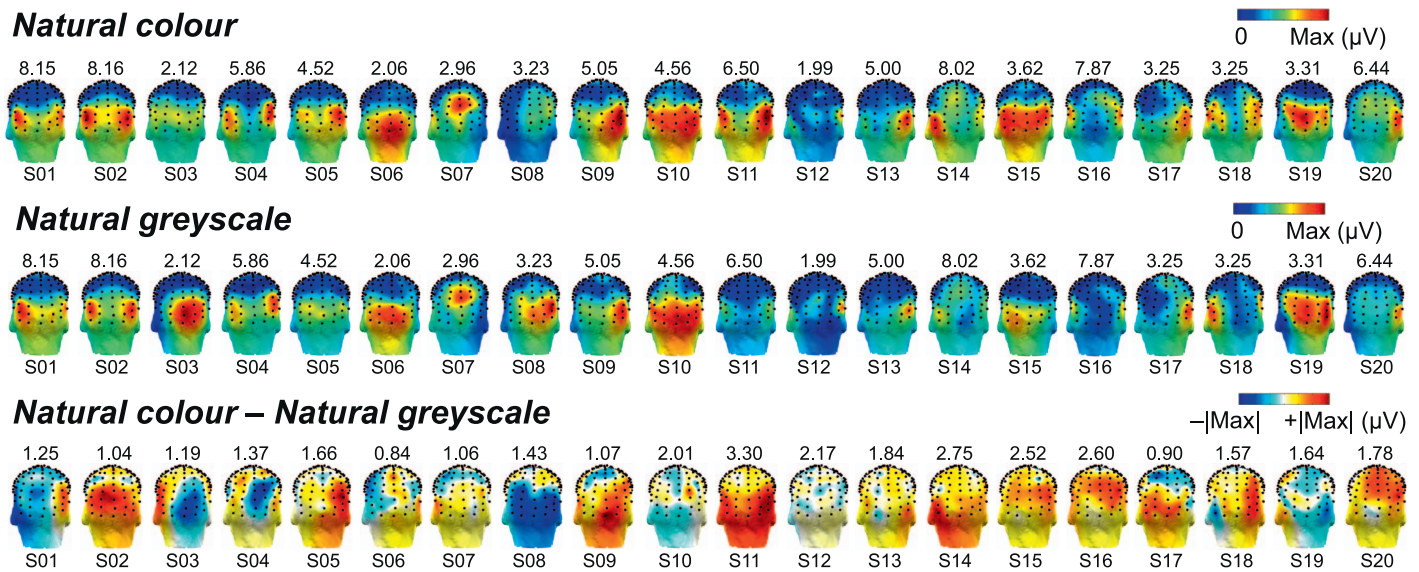


Figure 6. Experiment 2: Individual frequency-domain scalp topographies for the two natural image conditions and their differences. A back-of-the-head topography shows the sums of baseline-subtracted amplitudes across significant face stimulation harmonics (1.33–16.0 Hz, except 12 Hz) for each observer, none of whom participated in Experiment 1. To ease comparisons, the color scales are set to be the same across the two conditions within each observer, while the maximum amplitude (on top of each topography) varies across observers. For the difference topographies (natural color – natural grayscale), the color scales are also adapted to individual observers but the range was made symmetric around zero (i.e., $-|\text{maximum amplitude}|$ to $+|\text{maximum amplitude}|$, e.g., -1.25 to $+1.25$ μV for S01).

larger responses than natural grayscale images over bilateral occipitotemporal ROIs (IOT: 16.2% advantage, $p = 0.03$, rOT: 19% advantage, $p = 0.009$; see bar graph in Figure 5B), but no significant differences ($p > 0.11$) were found to support any color advantage in scrambled images. This suggests a mainly high-level color advantage for categorizing natural images, which was not found in the frequency-domain data in Experiment 1. For comparisons between natural and scrambled image conditions, all natural image conditions scored significantly larger responses than their corresponding scrambled image conditions regardless of ROI or the presence of color ($p < 0.001$; average response to scrambled images 2.8% of that to natural images), indicating a predominantly high-level response to physical structures in face categorization. For comparisons across ROIs, responses over rOT were significantly larger than those over IOT with natural images only (natural color: 21.9% advantage, $p = 0.005$, natural grayscale: 19% advantage, $p = 0.04$), but not with scrambled images ($p > 0.07$). Together, these data suggested a high-level, right-lateralized face categorization response that was enhanced in the presence of color when an orthogonal shape task was implemented.

Image stimulation frequency (12.0 Hz)

Average across all channels: Image stimulation responses were again found in all conditions (black lines

in Figure 5A). A z -score analysis showed that the first four harmonics (12–48 Hz) were significant (i.e., z score > 2.33) in all conditions. For consistency with Experiment 1, we decided to include also the fifth harmonic (60 Hz) when computing the harmonic-sum of baseline-subtracted amplitude for each observer and condition. It should be noted that including the fifth harmonic in subsequent analyses did not change the general results of the experiment, as the value of the nonsignificant baseline-subtracted amplitude was effectively close to zero.

Analyses similar to those in Experiment 1 were conducted here. A 2×2 repeated-measures ANOVA on channel-averaged, harmonic-summed, baseline-subtracted amplitudes showed significantly larger responses to scrambled images than to natural images by 16%, $F(1, 19) = 7.04$, $p = 0.02$. No significant differences were found between color and grayscale conditions, $F(1, 19) = 0.06$, $p = 0.81$, nor the interaction term, $F(1, 19) = 0.04$, $p = 0.85$. These results were consistent with those in Experiment 1.

Medial occipitoparietal area: The topographies (Figure 5C) again showed peak responses over the medial occipitoparietal area in all conditions. We defined a separate mOP₂ ROI based only on mean Experiment 2 data, which turned out to encompass channels Oz, Oiz, O2, POO6, and O1 (Figure 5D). This ROI overlapped partly with the mOP₁ ROI defined using Experiment 1 data (Figure 2D), though mOP₂ was shifted to a more inferior region. A 2×2 repeated-measures ANOVA on

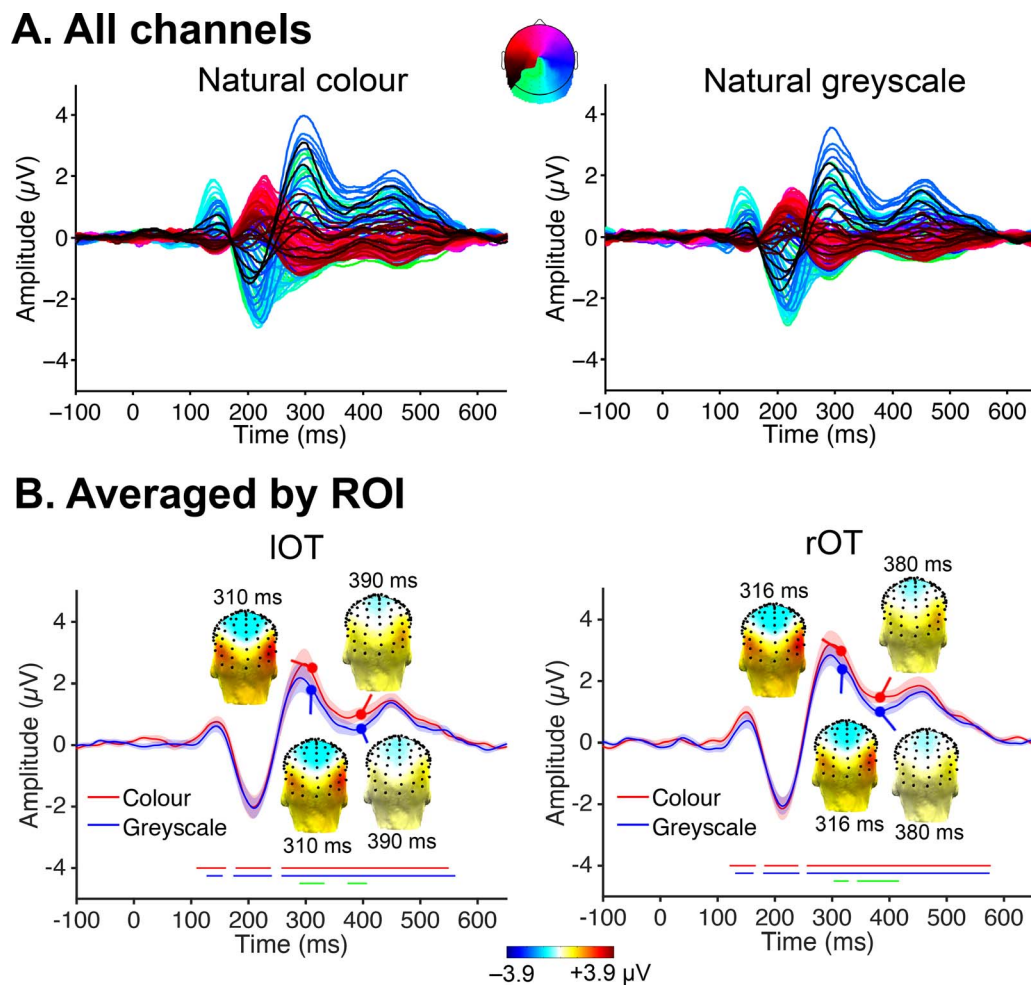


Figure 7. Experiment 2: Time-domain responses (following notch-filtering of 12 Hz and harmonics) to face stimulation in the natural image conditions (0 s: onset of face stimulation). (A) Waveforms for all 128 channels. The two-dimensional head map (viewed from top of the head) represents the color codes for the channels. (B) Waveforms averaged by ROI (see definitions in Figure 5D). Shaded areas represent ± 1 SEM across observers. For each ROI, the bottom horizontal bars represent significantly nonzero responses ($p < 0.05$), respectively, to natural color images (red), natural grayscale images (blue), and the difference between the two conditions (green) over 12 consecutive time points (i.e., 21.5 ms). At the midpoint of each green bar, the scalp topographies (back of the head) reveal superior face stimulation responses when the images contained color. The color scales are identical across all scalp topographies.

individual data averaged over the mOP₂ channels revealed similar pattern of results as the channel averages: scrambled > natural, 12% advantage, $F(1, 19) = 5.02$, $p = 0.04$; not significant for color vs. grayscale, $F(1, 19) = 0.85$, $p = 0.37$; also not significant for the interaction term, $F(1, 19) = 0.003$, $p = 0.95$. This again suggests a major contribution of the medial occipitoparietal region towards variations in image stimulation responses.

Time-domain analysis

Figure 7 shows the time-domain responses (after notch-filtering the image-stimulation rate response) for all 128 channels, and in particular, the two occipito-

temporal ROIs for the two natural image conditions. Similar to results in Experiment 1, we observed at least three distinctive components (P1-face, N1-face, P2/P3-face) time-locked to the periodic face stimuli, reflecting a face-selective process. All three face-selective components were significant ($p < 0.05$ over 21.5 ms) over both IOT and rOT regardless of the presence of image color (red and blue horizontal lines in Figure 7B). The time spans of the components (in the order of P1-face, N1-face, and P2/P3-face) were: IOT: color: 111–159 ms, 179–238 ms, and 259–548 ms; grayscale: 128–154 ms, 175–240 ms, and 259–560 ms; rOT: color: 122–165 ms, 183–240 ms, and 257–575 ms; and grayscale: 132–161 ms, 181–241 ms, and 257–574 ms. Pairwise t tests comparing ROI-averaged waveforms for natural color image and natural grayscale image conditions showed

an earlier, more pronounced (i.e., larger amplitude difference), and more sustained color advantage (green lines in Figure 7B) than in Experiment 1. In particular, significantly larger responses ($p < 0.05$ over 21.5 ms) to color images were found at 290–331 ms and 374–406 ms over IOT, and 304–327 ms and 345–415 ms over rOT. Note that the color advantage started from the peak of the P2/P3-face component, still a rather late latency relative to the onset of face stimulation (0 s).

General discussion

The current study evaluated the potential contribution of color to automatic face categorization in rapidly presented natural images by examining the responses to periodic face stimulation (1.33 Hz) against general object stimulation (12 Hz). In both experiments, as in previous studies with this paradigm (e.g., Retter & Rossion, 2016; Rossion et al., 2015), we obtained robust (i.e., significant for all participants) face categorization EEG responses to natural images, with the largest response found over the right occipitotemporal cortex. In addition, by comparing responses to natural images and those to their phase-scrambled, shape-deprived versions, we confirmed that processing associated to face categorization is mainly high level over the occipitotemporal areas, with right lateralization typically attributed to face-specific processing (e.g., Hécaen & Angelergues, 1962; Sergent, Ohta, & MacDonald, 1992; see Jonas et al., 2016 for intracerebral evidence with this paradigm). Importantly, the high-level characteristic face-selective responses obtained in this paradigm have been shown to differ quantitatively (i.e., much larger for faces) and qualitatively (i.e., in terms of spatiotemporal aspects) from other category-selective responses, such as those to body parts and houses (Jacques et al., 2016). Moreover, the periodic response is immune to potential temporal predictability arising from the periodicity of face presentations, as both temporally unpredictable appearances of faces and omissions of otherwise predictable face occurrences show comparable responses to temporally predictable faces (Quek & Rossion, 2017).

Here, while a minimal color advantage in face categorization found in Experiment 1 might be affected by an interaction with the task (detection of fixation-cross color change), modification to a more orthogonal task (Experiment 2: detection of fixation-cross shape change) led to a more pronounced, and more sustained color advantage, which we believe better reflects color's actual role in face categorization. In particular, we could not find evidence for image color to enhance neural face categorization responses until as late as 290 ms after onset of face stimulation, which coincides with

the beginning of the third face component (P2/P3-face). However, this color advantage extends to a little beyond 400 ms latency over the rOT region, save for a brief, intermittent break. The implications of such a late, but persistent enhancement are discussed as follows.

The late color advantage we found could be explained in terms of how much color contributes to face categorization relative to shape information. There is a view in object recognition research that, especially for objects with lower color diagnosticity, color's contribution steps up only when shape information becomes less diagnostic, for example, to categorization of animals that are structurally similar (Bramão, Faísca, Petersson, & Reis, 2012; Price & Humphreys, 1989; Tanaka & Presnell, 1999; Wurm, Legge, Isenberg, & Luebker, 1993). In the current experiments, faces were presented for 83 s (with sinusoidal contrast modulation), which, although brief, was demonstrably sufficient for consistent face detection in previous studies, even for grayscale images (see also Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Gegenfurtner & Rieger, 2000). Thus, participants' discrimination of faces from other objects may have already reached ceiling for grayscale images, masking potential early advantages of color during more challenging conditions, for instance, with shorter presentation durations, increased eccentricity, and added noise (e.g., Gegenfurtner & Rieger, 2000). That is, it could be argued that in the current experiments, the broadly similar three-dimensional physical structures across faces were sufficient to discriminate faces from other random objects that varied greatly in shape, diminishing a potential early contribution of color information. It is also worth noting that in our experiments, colored faces were contrasted with colored objects, which could also explain the lack of an early effect of color as sometimes reported in previous ERP studies, where effects of color could represent general enhanced activation, not specific to categorization per se (e.g., Zhu et al., 2013).

However, our results suggest that, at least for face categorization, color still contributes at a later stage. This is in fact consistent with reports of late, enhanced representation for nonface images presented in color (Gegenfurtner & Rieger, 2000; see also Yao & Einhäuser, 2008, reporting a color advantage for above-chance level categorization of animals from different-species, but not same-species, distractors during rapid serial visual presentation sequences). It is also consistent with previous findings that color information improves naming accuracy and shortens response time even during recognition of objects with lower color diagnosticity (Rossion & Pourtois, 2004), that image segmentation is performed better when color and texture cues are congruent rather than

conflicting (Saarela & Landy, 2012; though see Cant, Large, McCall, & Goodale, 2008), and that memory for shape can bias color perception in grayscale images (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006).

Our results thus generally agree with the “Shape + Surface” model of object recognition (Tanaka, Weiskopf, & Williams, 2001), where color plays a supporting role, with small but significant improvement, to primarily shape-driven face categorization (at least when stimuli are presented at 12 Hz), which is evident from the minimized responses to scrambled, shapeless images. The late emergence of the color advantage (beyond 290 ms after stimulus onset) is broadly consistent with the hypothesis that object shape and colour are processed in parallel and later combined (Tanaka et al., 2001). Indeed, a late color effect (N400 component) was also found for nonface object naming tasks when the objects are considered color diagnostic (Bramão, Francisco, et al., 2012). The hypothesis is also supported by a recent neuroimaging study showing that color-biased regions are segregated from face-selective regions along the ventral visual pathway in both humans and monkeys, but anterior color and shape areas show convergence (Laffer-Sousa, Conway, & Kanwisher, 2016). It is possible that color information facilitates the triggering of memory and knowledge of face and nonface objects for enhancing face categorization (Bramão, Francisco, et al., 2012). Such object color knowledge may even modulate color perception of the stimuli in low-level processing in early visual areas like V3 and V4 (Vandenbroucke, Fahrenfort, Meuwese, Scholte, & Lamme, 2016), though it should be noted that an early color effect (neither in P1-face nor N1-face) typically associated to image segmentation (e.g., Rossion et al., 2000) was not found in our study, again perhaps due to lack of face categorization difficulty for grayscale images as presented here.

We confirmed that processing associated to face categorization is mainly high level over the occipitotemporal areas, with right lateralization typically attributed to face-specific processing. This finding is also consistent with the aforementioned hypothesis that shape and color are processed in parallel and later combined, mainly during high-level face processing. Our data show that with these variable stimuli, the low-level contribution plays only a minor role in face categorization, as the average responses to phase-scrambled images (peaked over the low-level medial occipitoparietal area) account for only 4%–6% of those to natural stimuli in both experiments. This small contribution might be due to an undisturbed power spectrum after phase scrambling (Torralba & Oliva, 2003). Importantly also, we did not find color to have any modulating effect in the scrambled image conditions (Experiment 2), consistent with the hypothesis

that color facilitates later processing of face categorization.

Note that colored images necessarily contain additional chromatic contrasts (absent in grayscale images) that could potentially provide additional sensory inputs, resulting in a higher image stimulation response at 12 Hz and harmonics. However, we could not find such response difference between our natural color and grayscale conditions. One possibility is that any potentially added color inputs might be modulated by inhibitory effects from opponent chromatic contrasts between rapid successions of colored images. It is also possible that 12 Hz may not be an optimal stimulation frequency for capturing chromatic responses (Regan & Tyler, 1971).

Additionally, comparisons of the results from the two experiments allowed us to evaluate the influence of a distracting task on the potential effect of image colour to face categorization. We found a color advantage in the face stimulation responses to natural images in the frequency domain (1.33 Hz and harmonics) only in Experiment 2, where fixation shape change did not bias performance in the two natural image conditions. While this color advantage was absent in frequency-domain data in Experiment 1, probably due to distraction from fixation color changes, the time-domain analysis showed color advantage over the high-level, bilateral occipitotemporal regions at late onset latencies (beyond 376 ms), thus still showing some effect due to the presence of color. Using the shape task in Experiment 2, the time course showing the color advantage became more persistent between 290–415 ms and predominantly high level over the occipitotemporal regions. These results suggest that the contribution of color has a relatively late onset and is prone to be reduced by distractions. It is possible that the task demand reduces the effective use of color in face categorization (see also Zhu et al., 2013). This has practical implications on experimental designs on finding a small, but significant, effect like color in the current study.

Keywords: face categorization, color vision, natural scene, electroencephalography (EEG)

Acknowledgments

This work was supported by Nanyang Technological University Start-Up Grant and National Fund for Scientific Research (F.R.S.-FNRS, Belgium) postdoctoral fellowship to CO (FC 2773), Academic Research Fund (AcRF, Singapore) Tier 1 Grant 2018-T1-001-069 to CO and BR, F.R.S.-FNRS doctoral grant to TLR (FC 7159), and European Research Council (ERC) grant to BR (facesvpep 284025).

Commercial relationships: none.

Corresponding author: Charles C.-F. Or.

Email: charlesor@ntu.edu.sg.

Address: Division of Psychology, School of Social Sciences, Nanyang Technological University, Singapore.

References

- Bacon-Macé, N., Macé M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorization. *Vision Research*, *45*(11), 1459–1469. <https://doi.org/10.1016/j.visres.2005.01.004>
- Bindemann, M., & Burton, A. M. (2009). The role of color in human face detection. *Cognitive Science*, *33*(6), 1144–1156. <https://doi.org/10.1111/j.1551-6709.2009.01035.x>
- Boucart, M., Lenoble, Q., Quetelart, J., Szaffarczyk, S., Desprez, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision*, *16*(2):10, 1–13. <https://doi.org/10.1167/16.2.10>. [PubMed] [Article]
- Bramão, I., Faísca, L., Petersson, K. M., & Reis, A. (2012). The contribution of color to object recognition. In I. Kypraios (Ed.), *Advances in object recognition systems* (p. 73–88). <https://doi.org/10.5772/34821>
- Bramão, I., Francisco, A., Inácio, F., Faísca, L., Reis, A., & Petersson, K. M. (2012). Electrophysiological evidence for colour effects on the naming of colour diagnostic and noncolour diagnostic objects. *Visual Cognition*, *20*(10), 1164–1185. <https://doi.org/10.1080/13506285.2012.739215>
- Butterworth, S. (1930). On the theory of filter amplifiers. *Experimental Wireless & the Wireless Engineer*, *7*(85), 536–541.
- Cant, J. S., Large, M.-E., McCall, L., & Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception*, *37*(1), 57–78. <https://doi.org/10.1068/p5727>
- Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 660–675. <https://doi.org/10.1037/0096-1523.34.3.660>
- Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology*, *2*:342, 1–9. <https://doi.org/10.3389/fpsyg.2011.00342>
- De Dios, J. J. (2007). Skin color and feature-based segmentation for face localization. *Optical Engineering*, *46*(3), 037007. <https://doi.org/10.1117/1.2716016>
- De Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife*, *4*:e06564, 1–14. <https://doi.org/10.7554/elife.06564>
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research*, *40*(16), 2187–2200. [https://doi.org/10.1016/s0042-6989\(00\)00083-3](https://doi.org/10.1016/s0042-6989(00)00083-3)
- Dzhelyova, M., & Rossion, B. (2014). The effect of parametric stimulus size variation on individual face discrimination indexed by fast periodic visual stimulation. *BMC Neuroscience*, *15*:87, 1–12. <https://doi.org/10.1186/1471-2202-15-87>
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805–808. [https://doi.org/10.1016/S0960-9822\(00\)00563-7](https://doi.org/10.1016/S0960-9822(00)00563-7)
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*(6), 878–892. <https://doi.org/10.1080/13506280444000562>
- Graf, H. P., Chen, T., Petajan, E., & Cosatto, E. (1995). Locating faces and facial parts. In M. Bichsel (Ed.), *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition*, 41–46. Zurich, Switzerland: Universität Zürich. Multimedia Laboratory des Instituts für Informatik.
- Graf, H. P., Cosatto, E., Gibbon, D., Kocheisen, M., & Petajan, E. (1996). Multi-modal system for locating heads and faces. In M. E. Kavanagh (Ed.), *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 88–93. Los Alamitos, CA: IEEE Computer Society Press. <https://doi.org/10.1109/afgr.1996.557248>
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. <https://doi.org/10.1038/nn1794>
- Hécaen, H., & Angelergues, R. (1962). Agnosia for faces (prosopagnosia). *Archives of Neurology*, *7*(2), 92–100. <https://doi.org/10.1001/archneur.1962.04210020014002>
- Jacques, C., Retter, T. L., & Rossion, B. (2016). A single glance at natural face images generate larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-rele-

- vant categories in the human brain. *NeuroImage*, *137*, 21–33. <https://doi.org/10.1016/j.neuroimage.2016.04.045>
- Jonas, J., Jacques, C., Liu-Shuang, J., Brissart, H., Colnat-Coulbois, S., Maillard, L., & Rossion, B. (2016). A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proceedings of the National Academy of Sciences*, *113*(28), E4088–E4097. <https://doi.org/10.1073/pnas.1522033113>
- Jung, T.-P., Makeig, S., Lee, T.-W., McKeown, M. J., Brown, G., Bell, A. J., & Sejnowski, T. J. (2000). Independent component analysis of biomedical signals. *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation*, 633–644.
- Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *The Journal of Neuroscience*, *36*(5), 1682–1697. <https://doi.org/10.1523/jneurosci.3164-15.2016>
- Laganaro, M. (2014). ERP topographic analyses from concept to articulation in word production studies. *Frontiers in Psychology*, *5*:493, 1–10. <https://doi.org/10.3389/fpsyg.2014.00493>
- Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception*, *32*(8), 903–920. <https://doi.org/10.1068/p5007>
- Lewis, M. B., & Edmonds, A. J. (2005). Searching for faces in scrambled scenes. *Visual Cognition*, *12*(7), 1309–1336. <https://doi.org/10.1080/13506280444000535>
- Mollon, J. D. (1989). “Tho’ she kneel’d in that place where they grew...” The uses and origins of primate colour vision. *Journal of Experimental Biology*, *146*, 21–38.
- Mouraux, A., Iannetti, G. D., Colon, E., Nozaradan, S., Legrain, V., & Plaghki, L. (2011). Nociceptive steady-state evoked potentials elicited by rapid periodic thermal stimulation of cutaneous nociceptors. *Journal of Neuroscience*, *31*(16), 6079–6087. <https://doi.org/10.1523/jneurosci.3977-10.2011>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*(2), 176–210. <https://doi.org/10.1006/cogp.1999.0728>
- Otsuka, S., & Kawaguchi, J. (2009). Direct versus indirect processing changes the influence of color in natural scene categorization. *Attention, Perception & Psychophysics*, *71*(7), 1588–1597. <https://doi.org/10.3758/app.71.7.1588>
- Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *The Quarterly Journal of Experimental Psychology Section A*, *41*(4), 797–828. <https://doi.org/10.1080/14640748908402394>
- Quek, G. L., & Rossion, B. (2017). Category-selective human brain processes elicited in fast periodic visual stimulation streams are immune to temporal predictability. *Neuropsychologia*, *104*, 182–200. <https://doi.org/10.1016/j.neuropsychologia.2017.08.010>
- Regan, D., & Tyler, C. W. (1971). Wavelength-modulated light generator. *Vision Research*, *11*(1), 43–56. [https://doi.org/10.1016/0042-6989\(71\)90204-5](https://doi.org/10.1016/0042-6989(71)90204-5)
- Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, *91*, 9–28. <https://doi.org/10.1016/j.neuropsychologia.2016.07.028>
- Rossion, B., Alonso Pireto, E., Boremanse, A., Kuefner, D., & Van Belle, G. (2012). A steady-state visual evoked potential approach to individual face perception: Effect of inversion, contrast-reversal and temporal dynamics. *NeuroImage*, *63*(3), 1585–1600. <https://doi.org/10.1016/j.neuroimage.2012.08.033>
- Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, *11*(2):16, 1–21. <https://doi.org/10.1167/11.2.16>. [PubMed] [Article]
- Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297–1311. <https://doi.org/10.1016/j.visres.2011.04.003>
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (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*(1), 69–74.
- Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*,

- 1959–1979. <https://doi.org/10.1016/j.neuroimage.2007.10.011>
- Rossion, B., Jacques, C., & Jonas, J. (2018). Mapping face categorization in the human ventral occipito-temporal cortex with direct neural intracranial recordings. *Annals of the New York Academy of Sciences*, *1426*(1), 5–24. <https://doi.org/10.1111/nyas.13596>
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. <https://doi.org/10.1068/p5117>
- Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, *15*(1):18, 1–18, <https://doi.org/10.1167/15.1.18>. [PubMed] [Article]
- Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2007). Single-trial EEG dynamics of object and face visual processing. *NeuroImage*, *36*(3), 843–862. <https://doi.org/10.1016/j.neuroimage.2007.02.052>
- Saarela, T. P., & Landy, M. S. (2012). Combination of texture and color cues in visual segmentation. *Vision Research*, *58*, 59–67. <https://doi.org/10.1016/j.visres.2012.01.019>
- Sadr, J., & Sinha, P. (2004). Object recognition and Random Image Structure Evolution. *Cognitive Science*, *28*(2), 259–287. https://doi.org/10.1207/s15516709cog2802_7
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain*, *115*(1), 15–36. <https://doi.org/10.1093/brain/115.1.15>
- Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215. [https://doi.org/10.1016/s1364-6613\(00\)01626-0](https://doi.org/10.1016/s1364-6613(00)01626-0)
- Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153. <https://doi.org/10.3758/bf03207619>
- Torralla, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412. <https://doi.org/10.1088/0954-898x/14/3/302>
- Vandenbroucke, A. R. E., Fahrenfort, J. J., Meuwese, J. D. I., Scholte, H. S., & Lamme, V. A. F. (2014). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral Cortex*, *26*(4), 1401–1408. <https://doi.org/10.1093/cercor/bhu224>
- VanRullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research*, *46*(18), 3017–3027. <https://doi.org/10.1016/j.visres.2005.07.009>
- Wu, H., Chen, Q., & Yachida, M. (1999). Face detection from color images using a fuzzy pattern matching method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(6), 557–563. <https://doi.org/10.1109/34.771326>
- Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 899–911. <https://doi.org/10.1037//0096-1523.19.4.899>
- Yang, J., & Waibel, A. (1996). A real-time face tracker. In P. Storms (Ed.), *Proceedings of the Third IEEE Workshop on Applications of Computer Vision*, 1–6. Los Alamitos, CA: IEEE Computer Society Press. <https://doi.org/10.1109/acv.1996.572043>
- Yao, A. Y. J., & Einhäuser, W. (2008). Color aids late but not early stages of rapid natural scene recognition. *Journal of Vision*, *8*(16):12, 1–13, <https://doi.org/10.1167/8.16.12>. [PubMed] [Article]
- Zhu, W., Drewes, J., & Gegenfurtner, K. R. (2013). Animal detection in natural images: Effects of color and image database. *PLoS One*, *8*(10):e75816, 1–14. <https://doi.org/10.1371/journal.pone.0075816>