Cognition 225 (2022) 105172

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Face detection from patterns of shading and shadows: The role of overhead illumination in generating the familiar appearance of the human face

Colin J. Palmer [*], Erin Goddard, Colin W.G. Clifford

School of Psychology, UNSW Sydney, New South Wales 2052, Australia

A R T I C L E I N F O

Keywords:
Face detection
Illumination
Shading
Shadows
Light from above
Mooney faces

A B S T R A C T

Face detection in human vision relies on a stereotypical pattern of visual features common to different faces. How are these visual features generated in the environment? Here we investigate how characteristic patterns of shading and shadows that occur across the face act as a cue for face detection. We use 3D rendering to isolate facial shading under simulated lighting conditions, comparing the broad patterns of contrast that occur across the face when light arrives from different angles. We find that human performance in discriminating faces from non-face objects using these contrast patterns depends strongly on the lighting direction. In particular, light arriving from above the brow tends to facilitate face detection – consistent with the statistics of real-world lighting environments, in which light commonly arrives more strongly from above. Indeed, in a further experiment, we find that asymmetries in lighting that occur in complex and naturalistic lighting environments produce contrast patterns across the face that facilitate face detection. These effects occurred independent of the lighting direction relative to the viewer, suggesting that cues to face detection emerge from the interaction between face morphology and vertical asymmetries in lighting direction, independent of the viewer's knowledge or expectations about lighting direction. Comparison with the performance of an image classifier suggests that the effects of lighting direction partly reflect differences in image information that result from the interaction between shape and illumination, as well as face detection in human observers being better-tuned to the pattern of shading and shadows that occurs across an upright face that is lit from overhead.

The human face is one of the most familiar visual patterns that we encounter. Much can be gleaned from a person's face – who they are, their age, their emotions, and their focus of attention. Our visual system is specialized to extract this kind of information, but visual processing of faces first requires the ability to reliably detect faces in our environment. The challenge of face detection for both human- and machine-vision systems is that we all look different, and our appearance varies from moment-to-moment with changes in viewing angle, illumination, and expression (Adini, Moses, & Ullman, 1997). Face detection relies, therefore, on matching visual input against an internal template of features that are common to different faces, such as the basic spatial configuration of face features that can be captured in a simple contrast pattern (Fig. 1a; Johnson, 2005; Tsao & Livingstone, 2008). Critical features of this type appear to drive visual behaviour in newborns and adults (Farroni et al., 2005; Tomalski, Csibra, & Johnson, 2009), contribute to the tuning of 'face cells' in primate visual cortex (Ohayon, Freiwald, & Tsao, 2012; Tsao & Livingstone, 2008) and underlie perceptual phenomena like face pareidolia ('t Hart, Abresch, & Einhauser, 2011; Omer, Sapir, Hatuka, & Yovel, 2019). Similarly, in machine vision, algorithms have been developed that are able to perform face detection by exploiting a set of coarse intensity differences that appear relatively stable across images of faces (Sinha, 2002; Viola & Jones, 2001), such as a darker band across the eyes and lighter band across the upper cheeks. (For a review of the wider range of approaches to face detection used in computer vision, see Hasan, Ahsan, Abdullah-Al-Mamun, Newaz, & Lee, 2021).

What underlies the critical features of a face? One component is the *material properties* of the facial surface; for instance, coarse intensity differences around the eyes and mouth occur partly because the eyebrows, irises, and lips reflect light of a different intensity to the surrounding skin (Fig. 1b). Another component may be the *interaction between shape and lighting direction*. Shading and shadows are a near-ubiquitous feature of the faces we see in daily life. Under directional lighting, the recessed eye sockets, protruding nose, and prominent jaw of the human face contribute specific patterns of contrast across the facial surface. Are these patterns of shading and shadows a key feature of

* Corresponding author.
E-mail address: Colin.Palmer@unsw.edu.au (C.J. Palmer).

https://doi.org/10.1016/j.cognition.2022.105172
Received 22 August 2021; Received in revised form 11 May 2022; Accepted 12 May 2022
Available online 20 May 2022

what defines a face to the human visual system?

Pertinently, our visual system has evolved and developed in sensory environments with strongly directional lighting (e.g., Dror, Willsky, & Adelson, 2004). Light from the sun and sky arrives from above us, and indoor environments are commonly designed with light sources in the ceiling rather than the floor. The adaptation of our visual system to these conditions influences our basic experience of the world, including how we perceive the shape of shaded objects (Morgenstern, Murray, & Harris, 2011; Ramachandran, 1988). In the context of face perception, it has been recognized for some time that faces lit from above tend to appear more familiar to us – we can find it difficult to recognize people we know when we see them lit from below (Enns & Shore, 1997; Johnston, Hill, & Carman, 1992), and this is presumably why it is traditional to hold a flashlight below one's face when telling a scary story! The adaptation of our visual system to environments where light arrives more strongly from above may also play a role in face detection: Farroni et al. (2005) report that newborn babies have a preference to look towards faces lit from above rather than below, suggesting that from a very young age we may be "tuned to the particular distribution of dark and light patches characteristic of a face illuminated from above" (p.17248). There is also some evidence that the detectability of faces is modulated by lighting direction in adults (during interocular suppression; Stein, Peelen, & Sterzer, 2011), but the role of naturalistic illumination in providing cues that drive face detection in the human visual system is yet to be examined in detail.

A key phenomenon that is revealing about face detection is the perception of two-tone images or 'Mooney faces' – drawings or photographs of faces that are reduced to broad patterns of light and dark, but are often still recognizable as faces (Mooney, 1957). It is striking that, despite the amount of information that is abolished by this manipulation, enough critical features are seemingly retained, whether key contours or patterns of contrast, to match with an internal template of face structure (Cavanagh, 1991; Cavanagh & Leclerc, 1989; Moore & Cavanagh, 1998), highlighting the minimal nature of cues sufficient for face detection. As well as conveying a vivid impression of a face, two-tone images can elicit activation in the fusiform face area (Kanwisher, Tong, & Nakayama, 1998) and attract visual attention in newborns (Leo & Simion, 2009). However, the viewer's expectations about lighting direction may impact on their ability to detect faces in these images; Brodski, Paasch, Helbling, and Wibral (2015) report that two-tone images consistent with light arriving from *above the viewer* are more easily recognized as faces, including when a face *originally lit from below* is oriented upside down (hence, consistent with light arriving from above the viewer). This speaks to an 'expectation for light from above' affecting how readily we derive object shape from patterns of light and dark.

In the current study, we aim to test how simple visual features that enable face detection result from shading and shadows that occur across the face under directional lighting. We isolate patterns of luminance that arise specifically from the interaction between lighting direction and the morphology of internal face features, by rendering 3D face models with *uniform surface reflectance* and converting these to two-tone images that capture broad transitions between light and dark. These differ from two-tone images produced from photographs, where image structure depends partly on surface reflectance (e.g., the irises, eyebrows, skin, lips, and hair varying in the intensity of light they reflect) in addition to shading and shadows. Across three experiments, we test the ability of human observers to discriminate faces from unfamiliar non-face objects using contrast patterns produced by shading and shadows, and test whether performance depends on the angle that light falls across the face.

In considering the role of illumination in face detection, an important distinction can be made between how *lighting direction interacts with face shape* to generate a prototypical sensory pattern (e.g., shadows below the brows) and how the *viewer's expectations about lighting direction* aid in recovering 3D shape more generally. If there is an advantage in face detection for sensory patterns that are consistent with top-down lighting, this might occur either because (i) face detection relies on a template of visual features that is more closely tuned to the appearance of an upright human face under overhead lighting, or (ii) the visual system is generally better able to infer the shape of an object when the retinal image is consistent with light arriving from above the observer. The former puts emphasis on the lighting direction relative to the face, while the latter puts emphasis on the lighting direction relative to the observer. These frames of reference often align under ecological viewing conditions, but are dissociable for faces presented in a non-upright spatial orientation (Fig. 2).

In Experiments 1–3, we test whether face detection in human observers, when presented with contrast patterns produced by shading and shadows, depends on the lighting direction relative to the face. We manipulate the orientation of the images to test whether the pattern of

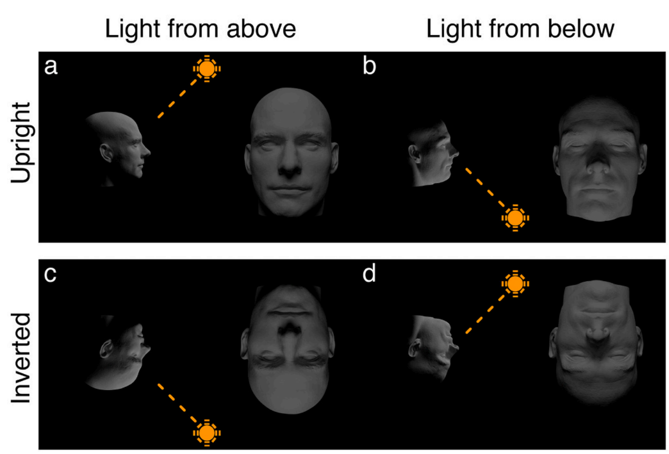

Fig. 2. The interaction between image orientation and lighting direction. (a) Faces are commonly seen in an upright orientation, with stronger lighting arriving from above the horizon. (b) When faces are lit from underneath, their appearance is altered by a different pattern of shading and shadows. (c, d) When the same images are presented upside-down, the lighting direction relative to the face is dissociable from the lighting direction relative to the observer. When describing our results, we use the terms 'light from above' and 'light from below' to refer to the lighting direction relative to the face, rather than the lighting direction relative to the observer.

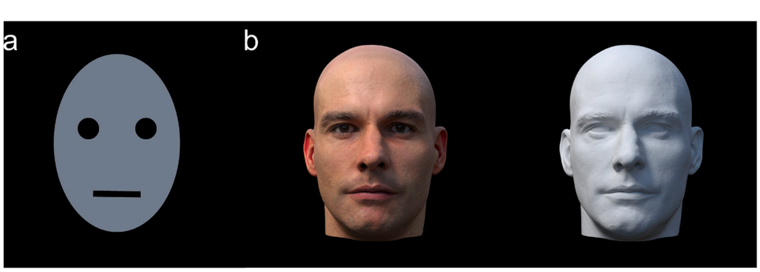

Fig. 1. (a) Face detection is enabled by sensory features that are common to different faces, such as a basic spatial pattern of brighter and darker regions. (b) These features are produced in part by the material properties of the facial surface; for example, the eyebrows, lips, skin, sclerae and irises reflect light of different intensity because they consist of different materials. However, when a 3D model of a face is rendered with uniform material properties (right), the pattern of shading and shadows that remains is still recognizable as a face. This pattern arises from the interaction between lighting direction and face morphology.

contrast produced by top-down lighting (relative to the face) provides an advantage even when inconsistent with light arriving from above the observer. Experiment 1 also tests whether face detection in contrast patterns produced by shading and shadows depends on the contrast polarity of the stimulus, and Experiment 2 tests whether performance depends on the intensity threshold used to define the pattern of contrast isolated in two-tone images. Experiment 3 examines how patterns of contrast produced across the face in complex and naturalistic lighting environments facilitate face detection. Finally, we report a set of image analyses that provide insight into the visual features associated with light falling on the face from above versus below and test whether the amount of low-level visual information available in the images differs with the direction of lighting (measured using an image classifier).

1. Experiment 1

1.1. Method

1.1.1. Participants

Data collection for all experiments was conducted online, with volunteers recruited via an online platform (Prolific; https://www.prolific.co/). Participants were eligible to participate in the study if they reported normal or corrected-to-normal vision, fluency in English, and UK nationality/residence. Participants were asked to use a desktop or laptop computer, and the browser Google Chrome. A minimum browser window resolution of 800 × 600 pixels was required to run the task. Each participant provided informed consent and was paid £8. The study was approved by the UNSW ethics committee. The sample for Experiment 1 included 37 adults with a mean age of 40 years (SD = 13 years) and gender split 23:14 (female:male). As we had no strong expectation about the size of effects under investigation, the target sample size (35–40) was chosen prior to data collection to fit within a range commonly used in behavioural research with a perceptual task.

1.1.2. Stimulus production

Face images were rendered under controlled lighting conditions in 3D graphics software, Blender 2.90 (The Blender Foundation, Amsterdam, The Netherlands). Image production was controlled using custom scripts in Python 3.5.3 (Python Core Team., 2017). Rendering was performed using Cycles, which is a physically-based rendering engine that simulates the path of light rays within a 3D environment and the interaction of this light with surfaces. In general, there are two key properties that determine the shading visible across the surface of an object: the angle that light arrives on the object (relative to the observer) and the 3D shape of the object. To ensure a realistic face shape, we used high-resolution models of facial geometry created by scanning real people. These models were produced by a 3D-scanning company, Ten24 (https://ten24.info/3d-scanning/) using a multi-camera array and photogrammetry. Models of six different people were used. The face models were rendered in Blender with Lambertian reflectance, uniform grey across the object surface. This meant that variations in intensity in the image arose specifically from the interaction between object shape and lighting direction (i.e., shading and cast shadows) rather than differences in reflectance across the face surface.

The rendering environment included a single light source. This was a 60 × 60 cm plane that emitted light towards the face, positioned at a distance of 1.5 m. The horizontal angle of the light source was either directly in front of the face, 45° left or 45° right. The vertical angle of the light source was 45° above or below the face. The faces were rendered from the perspective of a camera positioned 50 cm away, centered on the point in between the two eyes. The faces were oriented towards the camera, or rotated 30° left or right. The purpose of including different face models, horizontal orientations, and horizontal lighting directions was to increase variability in the image set and help prevent the participant forming strong expectations about the kind of face-pattern they were looking for. See Fig. S1 in Supplemental Material for examples of face renders produced under different simulated lighting directions.

A set of non-face objects were also created in Blender. The non-face objects were ellipsoids matched to the approximate height and width of the face models. The surface had the same uniform-grey, Lambertian reflectance as the face models. The 3D shape of the surface was modulated using a fractal Perlin noise texture that was mapped to the object surface and displaced the local geometry. This resulted in smooth modulation of the surface shape (i.e., protrusions and indentations) analogous to facial features. The object geometry was mirrored around the vertical midline. Overall, the renders of non-face objects shared a number of basic visual and structural properties with human faces, including symmetry around the vertical midline, larger-scale variations in shading that occur across an elliptical object under directional lighting, and finer-scale variations in shading due to smooth local modulations in surface shape. As there were six face identities, we created six non-face objects by using different noise patterns to modulate the geometry of the ellipsoid surface. The non-face objects were rendered under the identical conditions as the faces, including in vertical and horizontal lighting direction and object rotation. See Fig. S2 in Supplemental Material for examples of rendered non-face objects.

Further image processing was performed in MATLAB 2019b (Mathworks) to convert the grayscale renders to two-tone images (Fig. 3). The rendered images were low-pass filtered with a 2D Gaussian kernel (*SD* = 5 pixels; image size = 945 pixels square), which helps to produce clean transitions between darker and lighter regions after thresholding. The images were cropped with an elliptical mask to limit the visible area of the face to internal features. This meant that discriminating faces from non-face objects depended on internal face features, rather than the presence of a recognizable external contour formed by the ears, hair, and neck. To convert a grayscale image to a two-tone image requires selecting an intensity threshold that divides the image into darker and lighter regions. The choice of threshold is somewhat arbitrary, as there are different ways that one could define what the darker and lighter parts of an image are. The main concern in the current study was to use an objective and replicable method that could be applied in the same way across all images, leading to a fair comparison between images rendered under different lighting directions. We used an algorithmic method developed by Otsu (1979) that determines an optimal threshold based on the histogram of image grey-levels. This method uses a criterion that maximises the 'between-class variance' of pixels above and below the threshold, sensitive to both the mean grey level of pixels in each group and the number of pixels in each group. Greyscale images in Experiment 1 were encoded in an 8bit, non-linear intensity scale, and pixels within the elliptical crop mask were used when determining the threshold. The optimal threshold was determined separately for each image. To create a two-tone image, pixels of higher intensity than the threshold were set to white and all other pixels set to black. In Experiment 2, we test how results generalize across different methods for setting the threshold when creating two-tone images.

In addition to manipulating the direction of lighting used when rendering the face models, we also manipulated the image orientation and contrast polarity of the two-tone images. Image orientation was manipulated by rotating the image by 180°. Contrast polarity was manipulated by switching the black and white regions of the two-tone images.

1.1.3. Experimental task

Participants completed a face detection task, requiring discrimination between face and non-face stimuli (Fig. 4). The task was accessed from the participant's web browser and created using the JavaScript library jsPsych (de Leeuw, 2015) and JATOS (Lange, Kuhn, & Filevich, 2015). In each trial, a two-tone image was presented and participants judged whether the image depicted a human face or not. Each image was presented for approximately 100 ms. (For discussion of stimulus duration accuracy in jsPsych, see Kuroki, 2020). Responses were made by

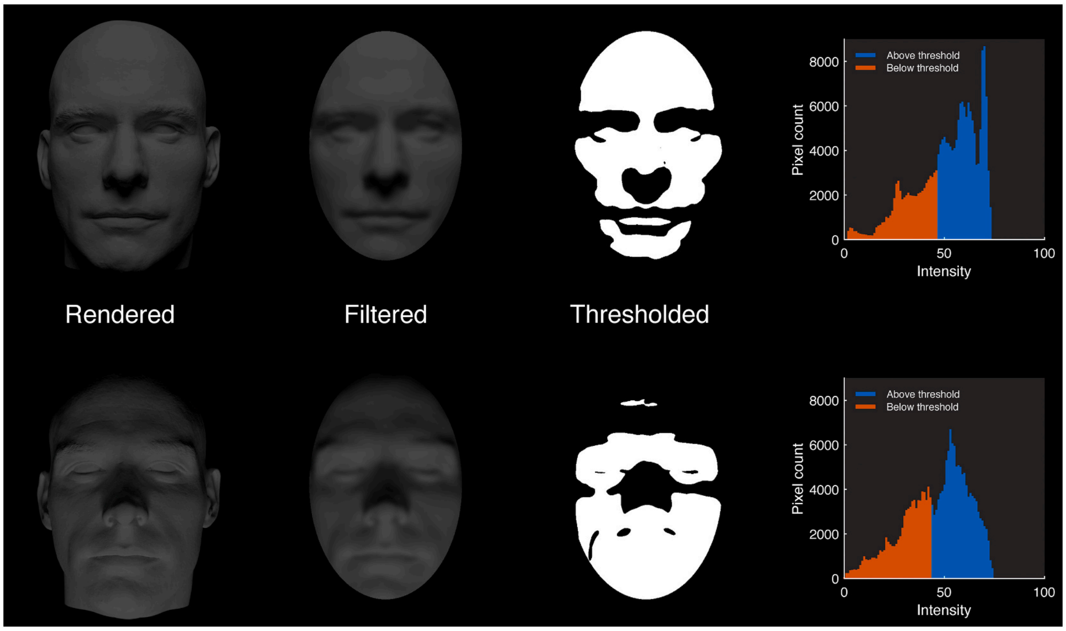

Fig. 3. Two-tone images produced under different, simulated lighting conditions. 3D-scanned models of human faces were rendered with a light source positioned 45° above the face (top row) or 45° below the face (bottom row). The face models were rendered with uniform reflectance to isolate the pattern of shading and shadows across the face. The renders were cropped and low-pass filtered, then converted to two-tone images using Otsu's method for threshold selection. This procedure isolated the broad transitions between lighter and darker regions caused by shading and shadowing across the internal features of the face.

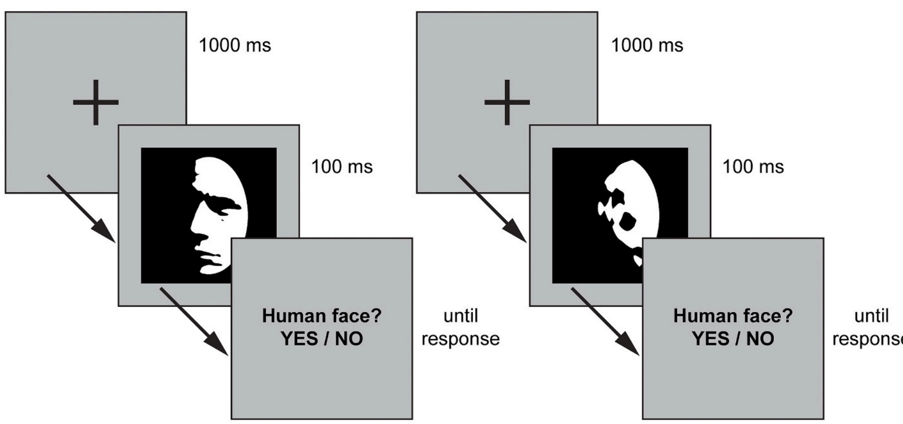

Fig. 4. Participants performed a face detection task, requiring discrimination between face and non-face stimuli. Two-tone images were presented briefly, depicting either a face or an abstract non-face object rendered under identical lighting conditions.

pressing one of two keys on the keyboard, with no time limit. Between trials, a fixation cross was presented for approximately 1000 ms. The image size varied randomly across trials, such that the region of the image within the elliptical crop mask was presented on screen 3.9–6.5 cm wide and 5.8–9.7 cm high. A calibration performed prior to the experiment helped to ensure that the images were presented at the same size across participants despite differences in monitor size. This involved each participant adjusting the size of an on-screen box to match the standard size of a credit card; this measurement of pixels-to-cm was used in the script to scale the presented images. Participants were instructed that the faces might be oriented normally or upside down. Trials were presented in a different random order for each participant, with all conditions interleaved. Participants completed 20 randomly-selected practice trials before beginning the task. The stimuli shown during practice trials were repeated in the main experiment.

1.1.4. Analysis

There were three independent variables used in the analysis. These were vertical lighting direction (light from above vs. below), image orientation (upright vs. upside down) and contrast polarity (natural vs. reversed). We compared face detection performance across these conditions using a three-way repeated-measures ANOVA. There were 108 trials for each condition, amounting to 864 trials in total. This included an equal number of faces and non-face objects, such that a measure of sensitivity in discriminating between faces and non-faces (d') could be calculated for each condition. It is useful to assess face detection performance in terms of sensitivity in discriminating faces from non-face objects, rather than accuracy in classifying the face images, as the latter could be driven by a general tendency to classify all images as faces. As described in the Stimulus Production section, trials within each condition varied in the identity of the face (or the identity of the non-face object), its horizontal rotation, and the horizontal lighting direction, but differences in performance across these stimulus dimensions were not analysed. Fig. S3 in Supplemental Material illustrates that the main effect of interest, the effect of vertical lighting condition, was apparent across head rotation and horizontal lighting conditions.

1.2. Results

1.2.1. Lighting direction

Face detection was strongly facilitated by light from above the face (Fig. 5). This was reflected in a main effect of lighting direction on sensitivity in discriminating faces from non-face objects, $F(1,36) = 202$, $p < .001$, $\eta_p^2 = 0.85$. For faces with natural contrast polarity, performance was better for images rendered with light arriving from above the brow of the face both when these images were presented upright, $t(36) = 15$, $p < .001$ (two-tailed), Cohen's $d = 2.4$, and when presented upside down, $t(36) = 12$, $p < .001$ (two-tailed), Cohen's $d = 2.0$. This indicates that the lighting direction *relative to the face* drives performance independent of the lighting direction *relative to the observer*. In other words, the particular pattern of shading and shadows that occurs when light arrives from above the brow (e.g., shadows within the eye sockets and below the nose) facilitates face detection regardless of whether this lighting direction is above or below from the perspective of the observer.

1.2.2. Image orientation

Face detection performance was consistently higher when images were presented upright rather than upside down, reflected in a main effect of image orientation, $F(1,36) = 37$, $p < .001$, $\eta_p^2 = 0.51$. The interaction between lighting direction and image orientation was not significant, $F(1,36) = 1.7$, $p = .20$, $\eta_p^2 = 0.05$, suggesting that the spatial inversion effect (i.e., better performance for faces presented upright compared to faces presented upside down) occurred similarly for faces lit from above and faces lit from below.

1.2.3. Contrast polarity

Face detection performance for images with reversed contrast polarity is shown in Fig. S4 in Supplemental Material. In the three-way ANOVA that included both natural- and reverse-polarity images, there was no significant main effect of polarity, $F(1,36) = 0.77$, $p = .39$, $\eta_p^2 = 0.02$. However, there was a significant interaction between polarity and lighting direction, $F(1,36) = 40$, $p < .001$, $\eta_p^2 = 0.53$. Post-hoc tests indicated that for images lit from above, reversing the polarity slightly reduced performance in discriminating faces from non-face objects, $t(36) = 4$, $p < .001$, mean difference in d' = −0.23. For images lit from below, reversing the polarity slightly *increased* performance, $t(36) = -2.5$, $p < .05$, mean difference in d' = 0.15. The three-way interaction between lighting direction, polarity, and image orientation was not significant, $F(1,36) = 0.69$, $p = .41$, $\eta_p^2 = 0.02$, suggesting that the interaction between polarity and lighting direction was similar for upright and upside-down images. Overall, these results suggest that the advantage for detecting faces that are lit from above derives partly from a familiar pattern of contrast polarity (e.g., darker regions in the eye sockets and below the nose). However, the effect of lighting direction on performance observed for natural-polarity images was also strongly apparent for polarity-reversed images. This occurred both when the polarity-reversed images were presented upright, $t(36) = 12$, $p < .001$ (two-tailed), Cohen's $d = 1.9$, and when presented upside down, $t(36) = 9$, $p < .001$ (two-tailed), Cohen's $d = 1.6$. This demonstrates that the spatial distribution of contrast differences contributes strongly to the advantage for detecting faces lit from above compared to faces lit from below, independent of contrast polarity.

There was also a significant interaction between polarity and image orientation, $F(1,36) = 16$, $p < .001$, $\eta_p^2 = 0.31$. Across the task, performance was lowest overall for images that were presented in both an unfamiliar orientation and reverse polarity (Fig. S4c).

2. Experiment 2

In Experiment 2, we test how different methods for deriving a contrast pattern from face images influence face detection performance. For faces with uniform reflectance, there is a straightforward relationship between lighting direction and image intensity: the brightest regions of the image occur where the surface of the face is angled more directly towards the light source, while the darkest regions of the image occur where the surface is angled more obliquely to the light source or receives cast shadows (Fig. 6a). One can create two-tone images where the pattern conveyed by black regions of the image captures only the darkest shadows that fall across the face or, alternatively, captures all but the very brightest surfaces of the face. Hence, thresholding the image at different intensities captures different types of contrast produced by the interaction of lighting direction with face morphology.

2.1. Method

2.1.1. Participants

The sample for Experiment 2 included 38 adults with a mean age of

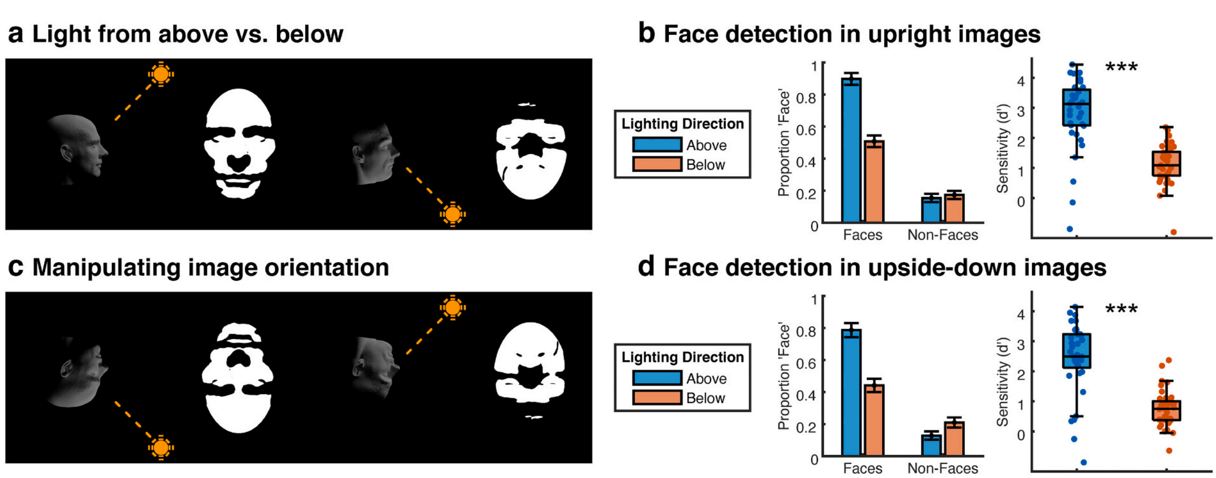

Fig. 5. The effect of lighting direction on face detection in images with natural contrast polarity. **(a)** Two-tone images produced from face models that were rendered in a 3D graphical environment with a light source 45° above or below the face. **(b)** Face detection was strongly facilitated when light arrived from above the face. Chance performance in discriminating faces from non-faces is indicated by a *d*' of zero. **(c)** When images are presented upside-down, both the face orientation and the direction of lighting are flipped relative to the observer. **(d)** Face detection in upside-down images was strongly facilitated when light arrive from 'above' the face, even though the image is consistent with light from below the observer. In these figures, lighting direction is labelled relative to the face rather than relative to the observer, such that the images shown for each condition in panel d are the same images as in the corresponding condition of panel b, but flipped upside down. Barplots show the mean ± 1 standard error. Boxplots show the median, interquartile range, and full range of the data excluding scores that lie beyond the limits of the box by more than 1.5 times the interquartile range. *** $p < .001$.

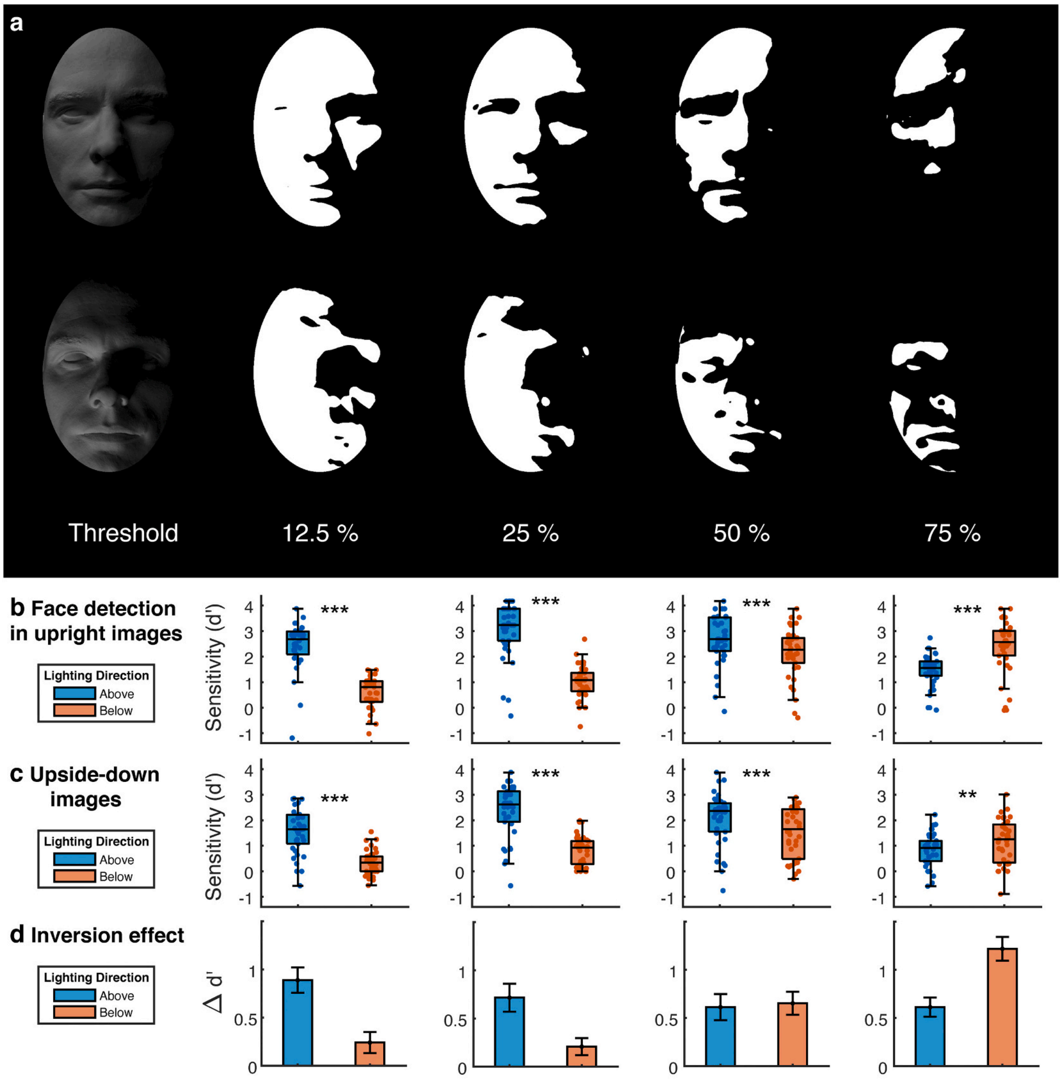

Fig. 6. Does face detection in simple contrast patterns depend on the intensity threshold that divides light and dark regions? **(a)** Each face render was thresholded at four different levels, expressed here as a percentage of the average maximum intensity across images. At the highest threshold level, the white regions of the two-tone image correspond only to the very brightest regions of the face (the regions of the face surface that are angled most directly towards the position of the light source in the 3D graphical environment). The top row shows a face lit from the upper-left, and the bottom shows a face lit from the lower-left. **(b)** Face detection tended to be higher for two-tone faces lit from above, but this trend reversed for the highest threshold level. **(c)** When the images were presented to participants upside down, the difference in performance between lighting conditions was in the same direction as for the upright images, indicating that the lighting direction relative to the face drives performance independent of the lighting direction relative to the observer. **(d)** The spatial inversion effect (i.e., difference in performance between upright versus upside-down images) differed across lighting directions and threshold levels. Barplots show the mean ± 1 standard error. Boxplots show the median, interquartile range, and full range of the data excluding scores that lie beyond the limits of the box by more than 1.5 times the interquartile range. *** $p < .001$; ** $p < .005$.

39 years (SD = 15 years) and gender split 23:15 (female:male). There was no overlap with the sample from Experiment 1. In Experiment 1, we observed a strong effect size for the difference in performance between images lit from above versus below. A power analysis indicated that only a small sample ($n = 5$) would be necessary to achieve >95% power in detecting an effect of this magnitude. However, we chose to maintain a similar sample size to facilitate comparison across experiments.

2.1.2. Stimulus production

Stimuli were produced using similar methods as Experiment 1, but with a parametric manipulation of the threshold used to convert grayscale renders into two-tone images (Fig. 6a). Because the greyscale images were produced in the same rendering environment, a given image grey level corresponded to the same physical intensity of light across all images. This meant that we could apply an intensity threshold globally across images, dividing the darker and lighter regions of each image based on a specific physical intensity, and vary this threshold parametrically. For this purpose, in Experiment 2, greyscale intensities were encoded in a linear, 32-bit format, then thresholded at four different intensity levels. The thresholds were set at 12.5%, 25%, 50%, and 75% of the average maximum intensity across images. This approach complements Experiment 1, in which the threshold used to produce the two-tone images was optimised on a per-image basis using Otsu's method rather than applied globally across images.

2.1.3. Experimental task and analysis

Participants completed the same face detection task as in Experiment 1, but with different images presented. There were three independent variables used in the analysis. This included vertical lighting direction (light from above vs. below), image orientation (upright vs. upside down) and threshold level (12.5% vs 25% vs 50% vs 75% of the average maximum intensity). We compared face detection performance across these conditions using a repeated-measures ANOVA. There were 54 trials for each condition, amounting to 864 trials in total. These included an equal number of faces and non-face images. Trials in each condition also varied in the identity of the face (or non-face object), its horizontal rotation, and the horizontal lighting direction. Compared to Experiment 1, half the number of face and non-face identities were used to maintain the same total number of trials.

2.2. Results

2.2.1. Lighting direction and threshold level

Consistent with the results of Experiment 1, face detection was strongly facilitated by light from above the face. This was reflected in a main effect of lighting direction on sensitivity, $F(1,37) = 179$, $p < .001$, $\eta_p^2 = 0.83$. However, the effect of lighting direction depended on the threshold level, with a significant interaction between these factors, $F(3,111) = 138$, $p < .001$, $\eta_p^2 = 0.79$. In particular, sensitivity was higher for top-lit faces compared to bottom-lit faces for the threshold levels of 12.5%, 25%, and 50%, but higher for bottom-lit faces for the 75% threshold level (Fig. 6b). Paired samples *t*-tests indicated that sensitivity was significantly different between lighting conditions for each threshold level. This was the case for both upright images ($p < .001$ for all; Cohen's $d = 2.29, 2.10, 0.64, -1.21$, respectively, across threshold levels) and upside-down images ($p < .005$ for all; Cohen's $d = 1.53, 1.88, 0.79, -0.54$). The reversal of performance trends across lighting directions for two-tone images produced by thresholding at a high intensity level (75%) compared to those produced by thresholding at lower intensity levels (12.5%, 25%, 50%) suggests a difference in information useful for face detection that is carried in the pattern formed by the very brightest regions of the original face images compared to the information carried by the broader contrast between lighter and darker regions of the faces. (This point is discussed further in the Discussion section *Dependence on Intensity-Threshold Level*). For each threshold level, the lighting condition with the highest sensitivity in the upright images (i.e., top-lighting for threshold levels 12.5%, 25%, and 50%, and bottom-lighting for threshold level 75%) also had the highest sensitivity for upside-down images (Fig. 6b and c). Thus, consistent with Experiment 1, the direction of lighting relative to the face determined performance, independent of the direction of lighting relative to the observer.

2.2.2. Image orientation

Face detection performance was consistently higher when images were presented upright rather than upside down, reflected in a main effect of image orientation, $F(1,37) = 81$, $p < .001$, $\eta_p^2 = 0.69$. However, the strength of the spatial inversion effect (i.e., the difference in performance between images presented upright versus upside down) depended on both the lighting direction and threshold level (Fig. 6d). This was reflected in a significant interaction between image orientation and threshold level, $F(3,111) = 8.3$, $p < .001$, $\eta_p^2 = 0.18$, and a three-way interaction between image orientation, threshold level, and lighting direction, $F(3,111) = 19$, $p < .001$, $\eta_p^2 = 0.34$. In Fig. 6d, one can see that the spatial inversion effect tended to be greater for top-lit faces compared to bottom-lit faces at lower threshold levels, and greater for bottom-lit faces compared to top-lit faces at the highest threshold level. Overall, the tendency for spatial inversion to reduce performance suggests that the mechanisms facilitating face detection for top-lit faces (at lower thresholds) and bottom-lit faces (at higher thresholds) are sensitive in both cases to the typical upright configuration of a face.

3. Experiment 3

In experiments 1 and 2, faces were rendered in a 3D-graphical lighting environment that contained a single light source angled 45° above or below the face. In real-world environments, illumination typically arrives on an object from all directions at once, due to the presence of multiple primary light sources, extended light sources (e.g., the sky), and inter-reflections from other surfaces present in the environment. In Experiment 3, we test whether shading patterns that occur across a face in more complex and naturalistic simulated lighting environments facilitate face detection.

3.1. Method

3.1.1. Participants

The sample for Experiment 3 included 40 adults with a mean age of 47 years (SD = 13 years) and gender split 22:18 (female:male). There was no overlap with the samples from experiments 1 or 2. Data from three additional participants were excluded from the analysis because these participants reported technical difficulties in completing the online task. The sample size was chosen to match experiments 1 and 2.

3.1.2. Stimulus production

Complex lighting environments were simulated using a set of high-dynamic range illumination (HDRI) maps captured from real-world environments. These are panoramic images that represent the intensity of light arriving at a particular point in space from all directions, i.e., spanning 360° horizontally and 180° vertically. These are projected onto a rectangular surface for display purposes in Fig. 7a, and were used as spherical illumination environments in Blender. Six different HDRI maps were used, captured from environments with both natural and artificial light sources, and at different times of day. This included a forest, an open field on a sunny day and in the evening, a residential garage, an apartment building corridor, and an urban street. The HDRI maps were produced by aifosDesign (https://www.aifosdesign.se/kategori/hdri/).

The HDRI maps tended to have stronger illumination arriving from angles above the horizon compared to angles below the horizon (Fig. S5 in Supplemental Material). This is consistent with the general observation that light tends to come more strongly from above in real-world environments (e.g., Dror et al., 2004). Faces were rendered with the illumination environment either in a natural vertical orientation, with the sky or ceiling above the face model and camera, or with the vertical axis inverted, such that the sky or ceiling was below the face model and camera (Fig. 7b). The horizontal orientation of the HDRI maps was set such that a dominant light source present in the environment (e.g., the sun or a nearby artificial light source) was angled 30° to the left or right of the face.

Greyscale images of the face and non-face models were rendered in Blender in a linear, 32-bit format, then low-pass filtered and cropped in MATLAB as described for Experiment 1. Real-world environments vary dramatically in their absolute level of luminance (e.g., a sunny day vs. a dim indoor environment), which we accounted for by normalizing each image. This was done by taking pixels from the face region visible within the elliptical crop mask, subtracting their mean intensity and dividing by 3 standard deviations. This results in the relevant pixels in each image having a mean intensity of 0 and standard deviation of one-third. The images were then scaled and clipped to an 8-bit range, and a threshold was determined from the grey-level histogram for each image using the same method as Experiment 1. The resulting two-tone images thus brought out contrast between darker and lighter regions of the face independent of the absolute mean intensity. This is based on the logic that the human visual system will typically adjust to the prevailing illumination conditions, bringing out the contrast that exists within the present environment.

3.1.3. Experimental task and analysis

Participants completed the same face detection task as in the two previous experiments, but with different images presented. There were two independent variables used in the analysis. These were the vertical orientation of the lighting environment (natural vs. inverted) and image orientation (upright vs. upside down). We compared performance across

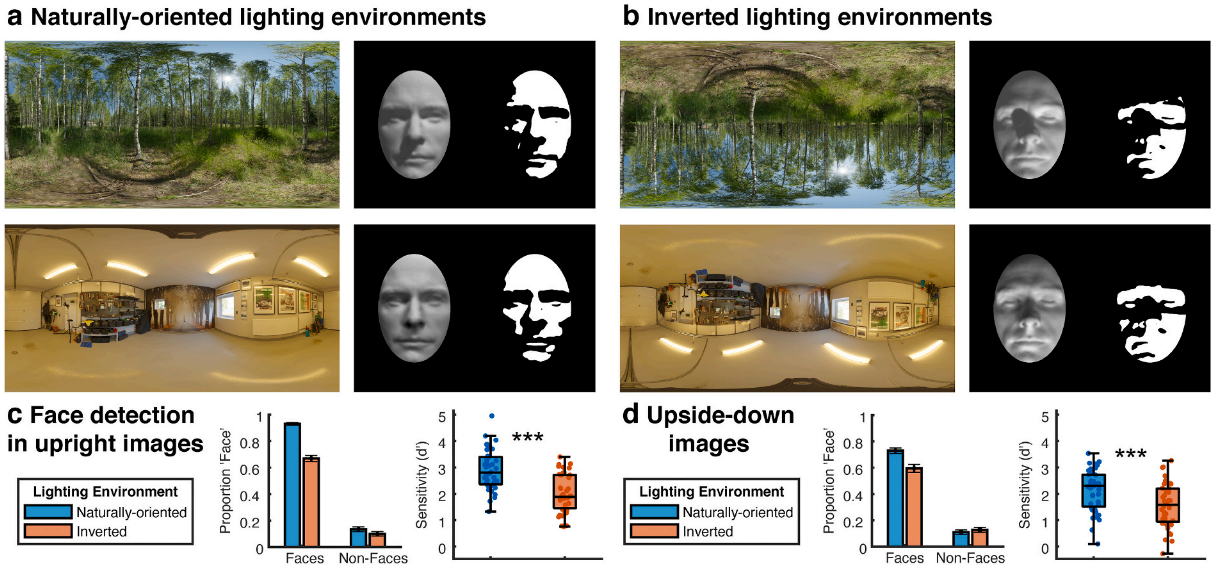

Fig. 7. Facial shading patterns produced in complex lighting environments that were simulated using 3D graphical rendering. **(a)** Face models were rendered using high-dynamic range illumination maps recorded from a number of different real-world environments. Thresholding the face images isolated the broad transitions between lighter and darker regions. **(b)** Faces were also rendered in lighting environments that were inverted along the vertical axis, i.e., with the sky or ceiling below the face. **(c)** Face detection performance was higher for contrast patterns produced in naturally-oriented lighting environments. **(d)** When the same contrast patterns were presented to participants upside down, face detection performance was higher for contrast patterns consistent with lighting environments oriented naturally *relative to the face*, even though this is upside down *relative to the observer*. Barplots show the mean ± 1 standard error. Boxplots show the median, interquartile range, and full range of the data excluding scores that lie beyond the limits of the box by more than 1.5 times the interquartile range. *** $p < .001$.

these conditions using a repeated-measures ANOVA. There were 216 trials for each of these conditions, resulting in 864 trials in total. This included an equal number of faces and non-face objects. Trials in each condition also varied in the identity of the face (or non-face object), its horizontal rotation, and the horizontal orientation of the lighting environment.

3.2. Results

3.2.1. Vertical orientation of the lighting environment

Face detection was strongly facilitated by a natural orientation of the lighting environment relative to the face (Fig. 7c and d). This was reflected in a main effect of lighting orientation on sensitivity in discriminating faces from non-face objects, $F(1,39) = 176$, $p < .001$, $\eta_p^2 = 0.82$. Performance was better for images rendered with a natural orientation of the lighting environment both when these images were presented upright, $t(39) = 11.9$, $p < .001$ (two-tailed), Cohen's $d = 1.88$, and when presented upside down, $t(39) = 8.6$, $p < .001$ (two-tailed), Cohen's $d = 1.36$. Thus, consistent with experiments 1 and 2, the direction of lighting relative to the face determined performance, independent of the direction of lighting relative to the observer.

3.2.2. Image orientation

Face detection performance was better when images were presented upright rather than upside down, reflected in a main effect of image orientation, $F(1,39) = 62$, $p < .001$, $\eta_p^2 = 0.62$. There was also a significant interaction between the orientation of the lighting environment and the image orientation, $F(1,39) = 15$, $p < .001$, $\eta_p^2 = 0.29$. The spatial inversion effect (i.e., the difference in performance between images presented upright versus upside down) was stronger for images consistent with a natural orientation of the lighting environment relative to the face compared to an inverted lighting environment, $t(39) = 4$, $p < .001$, Cohen's $d = 0.63$, mean difference = 0.35 (SE difference = 0.09).

4. Image analysis

Distinct patterns of contrast and edges occur when face models are rendered under lighting that arrives more strongly from above versus below. These are illustrated in Fig. 8, averaged across the two-tone images used in each experiment. There is a clear qualitative distinction between the patterns of shading and shadows that occur across the face under different lighting conditions, with darker regions within the eye sockets and below the nose tending to occur when faces are lit from above, even when averaging across images that differ in head direction and horizontal lighting direction, and even for images that are consistent with complex simulated lighting environments in which light arrives from all directions at once (but with naturally-occurring asymmetries in the magnitude of illumination across vertical and horizontal angles).

Fig. 9a shows the same image analysis for face stimuli presented in Experiment 2 when separated by threshold level. The 'reversal' in performance trends across lighting conditions that we observed in Experiment 2 (for two-tone images thresholded at a high intensity level) provided a further opportunity to visualize the image features that vary with human performance. We focused on the 25% and 75% threshold conditions, as peak performance occurred in these conditions for images consistent with lighting from above and images consistent with lighting from below, respectively (Fig. 6b). In particular, for each lighting condition, we subtracted the edge density map computed from images in the lower-performance condition from the edge density map computed from images in the higher-performance condition (Fig. 9b). The resulting map of edge density *differences* indicates the regions of the two-tone images in which edges tend to occur when performance is better compared to when performance is poorer, independently for each lighting condition. There is a notable similarity between the features associated with an advantage in face detection for the light-from-above condition and the light-from-below condition – in both lighting conditions, better performance is associated with edges occurring at the level of the eyes, nostrils, and lips (Fig. 9b).

To test the extent to which low-level image differences might contribute to categorization performance across the three experiments, we used a classification analysis to measure whether the face and non-face images were more distinguishable in some conditions than others. Classifier accuracy provides a measure of how reliably the face and non-face images can be categorized in each condition (e.g., when lit from

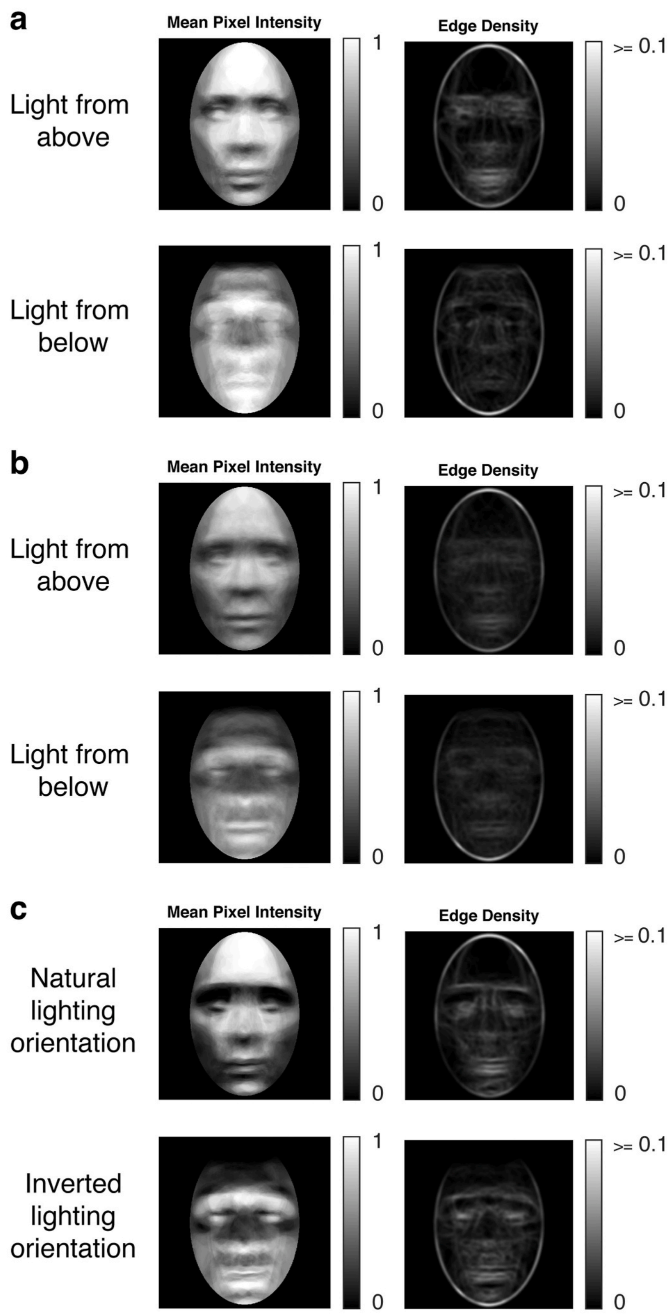

Fig. 8. Visualising image structure of two-tone faces presented in Experiment 1 (a), Experiment 2 (b) and Experiment 3 (c). For each experiment, the images in the left-hand column show mean pixel intensity for upright faces with natural contrast polarity, averaged across identities, head rotation, and horizontal lighting direction. These images illustrate how the pattern of contrast across the face produced by shading and shadows tends to differ when the face is lit from above versus below (Experiment 1–2) or when rendered in naturally-oriented versus inverted lighting environments (Experiment 3). The images in the right-hand columns are heatmaps of where edges occur in the images. Edges were identified using the Sobel method implemented in MATLAB, for upright faces with natural contrast polarity. The heatmaps show the proportion of images containing an edge, after spatial smoothing with a 2D Gaussian filter ($SD = 5$ pixels).

above vs below) based on low-level image information.

4.1. Method

To approximate the information in each image that would be accessible to the early visual system, we first filtered each image using the early levels of the 'HMAX' model of object recognition (Riesenhuber & Poggio, 1999; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007). Specifically, we used the output of the 'C2' or 'complex composite' layer of the HMAX model, with response properties designed to approximate those of some V4 neurons (Serre et al., 2007). This is the fourth layer of the model, preceded by S1, C1, and S2 layers. We used the Matlab implementation of this model available from https://maxlab.neuro.georgetown.edu/hmax.html (accessed March 2021). Each stimulus image (face and non-face objects) was 945 pixels square. The first layer of the HMAX model (S1) uses a bank of orientation-tuned filters: we defined filters at each of 4 equally spaced orientations (horizontal, vertical and +/− 45 degrees from vertical), and each of 19 receptive field sizes (23 to 131 pixels across, in 6 pixel steps). Each filter was odd-symmetric, with spatial frequency that scaled with size so that each included 1.5 cycles of a sinusoidal modulation. For remaining model parameters we used the default values, including a C1 layer with 8 scale bands (1 to 17 in steps of 2), that had spatial pooling ranges of 8 to 22 (in steps of 2), and the universal set of patches supplied with the Matlab HMAX model. Using these universal patches yields 400 responses at each of 8 scales, for a total of 3200 C2-layer responses per image. We generated HMAX C2-layer responses for each image used in the experiments with human observers. For each experiment, we reduced the dimensionality of the dataset of model responses (originally 3200 dimensions) by applying principal components analysis (PCA), implemented using Matlab function *pca*, and retaining the scores from the first n components that accounted for 99% of the variance across images (Experiment 1: $n = 757$; Experiment 2: $n = 749$; Experiment 3: $n = 741$).

We trained classifiers on a series of face vs non-face discriminations using the PCA scores of HMAX model responses. In every case, we used a naïve Bayes classification, implemented using the Matlab functions '*fitcdiscr*' and '*predict*' with variable '*type*' set to '*diaglinear*'. We repeated these classification analyses with alternative discriminant functions ('*linear*' and '*diagQuadratic*') and obtained qualitatively similar results (data not shown), suggesting that classifier performance reflects differences in image information rather than being specific to a particular classification method. We were particularly interested in whether images of different lighting direction and/or threshold level varied in the availability of low-level information that could be used to distinguish face from non-face images, so we confined classification analyses to images with upright orientation and natural contrast polarity. For each experiment, the total set of images was partitioned according to identity (6 face and 6 non-face identities for Experiment 1; 3 face and 3 non-face identities for Experiments 2 and 3; 864 images in total in each experiment). We trained classifiers to discriminate face from non-face images using all except one pair of identities (one face and one non-face object), then we evaluated the classifier on how accurately it could predict the category of the held-out images. We repeated this procedure several times, so that each image was included once in the test data set, and report the classifier sensitivity in discriminating face and non-face images across runs. By training the classifiers on the output of the HMAX model we were seeking to reduce the likelihood that classifier accuracy could be based on some spurious feature that was unlikely to contribute to human performance. For example, if we instead gave the classifier the intensity of each pixel in the image it would be possible that the classifier could learn that the presence of white at a particular pixel was predictive of face vs non-face, but this feature would be very unlikely to drive human performance. Similarly, by always testing classifier performance on held out pairs of identities, we reduced the chance that robust classifier performance could be based on very specific local features that were unique to the particular identities in the training set.

To test for statistical significance, we used bootstrapping to generate a null distribution of classifier performance. We divided the images into pairs of identities (each pair consisting of one face and one non-face object). For each bootstrapped dataset, we generated a training data-set by randomly shuffling the face/non-face category labels within each

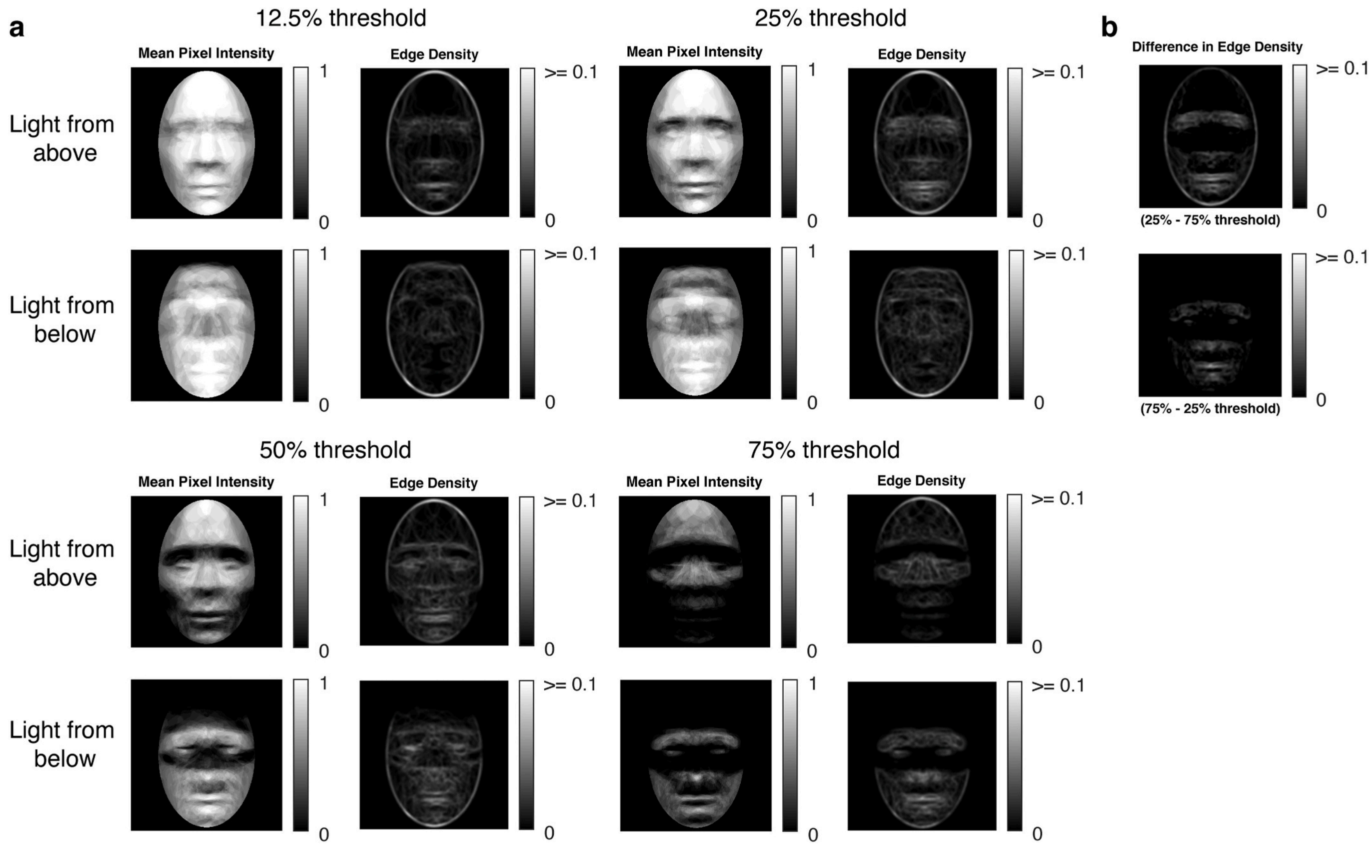

Fig. 9. Visualising image structure of two-tone faces presented in Experiment 2, separated by threshold level. (a) Face images were converted to two-tone images by thresholding at one of four intensity levels, expressed here as a percentage of the average maximum intensity across images. For each threshold level, the images in the left-hand column show mean pixel intensity of the two-tone images generated from upright faces, averaged across identities, head rotation, and horizontal lighting direction. The images in the right-hand columns are heatmaps of where edges occur in the images. (b) To help visualize the image features that support face detection, edge density was compared between conditions that varied in behavioural performance (Fig. 6b). In particular, the upper panel shows the difference in edge density between higher- and lower-performance image sets consistent with *light from above* (25%–75% threshold conditions), and the lower panel shows the difference in edge density between higher- and lower-performance image sets consistent with *light from below* (75%–25% threshold conditions). These plots illustrate that in both lighting conditions, better performance is associated with edges occurring at the level of the eyes, nostrils, and lips.

pair of identities. We repeated our classification analysis as for the original analysis, except that we used the shuffled-label data as training data, and the original-label data to test classifier performance. We repeated this process 10,000 times, yielding a null distribution of classifier performance for each case, used to define 95% confidence intervals. We also used the results of these bootstrapping analyses to generate null distributions of 10,000 difference values (lighting angle from above versus below) for each comparison. We report *p*-values for each comparison based on the proportion of these null absolute difference values that exceeded the observed absolute difference (two-sided tests).

4.2. Results

Classifier performance is shown in Fig. 10. In the image sets for all three experiments, the ability of the classifier to learn to distinguish faces from non-face objects differed depending on the direction of lighting. In experiments 1 and 2, human performance (reported in Figs. 5 and 6) trended in the same direction as classifier performance across lighting conditions. This included the advantage for lighting from above observed at the lowest threshold level in Experiment 2 trending in the reverse direction at the highest threshold level. This suggests that human performance in these experiments might be accounted for partly by differences in low-level information available in the two-tone faces lit from above versus below. However, for the images used in Experiment 3, classifier performance across lighting conditions differed in the opposite direction to that found for human observers. In other words, humans performed better in detecting faces in images consistent with light arriving more strongly from above (rather than below) even when there was less low-level information available in the former. This suggests that prior familiarity with faces lit from above plays a role in determining human performance.

5. Discussion

There is a characteristic form to the human face – to put it unflatteringly, we have sunken eye sockets, a protruding nose, and pouty lips. These features of the face produce distinctive patterns of shading and shadows under directional lighting, patterns that we are likely accustomed to seeing even if not explicitly aware of them. In the present study, we isolated the broad patterns of contrast that occur across faces due to shading and shadowing effects, by 3D-rendering face models with uniform reflectance under different simulated lighting conditions, then thresholding the images. Human observers were able to detect faces in these contrast patterns, including those consistent with both simple and complex lighting environments. Performance was sensitive to the typical upright configuration of the face and depended partly on natural contrast polarity but moreso on the spatial distribution of contrast independent of polarity. Perhaps most interestingly, face detection performance was strongly facilitated by contrast patterns consistent with light arriving from above the brow (with one exception, discussed below). This advantage for light from above was driven by how lighting

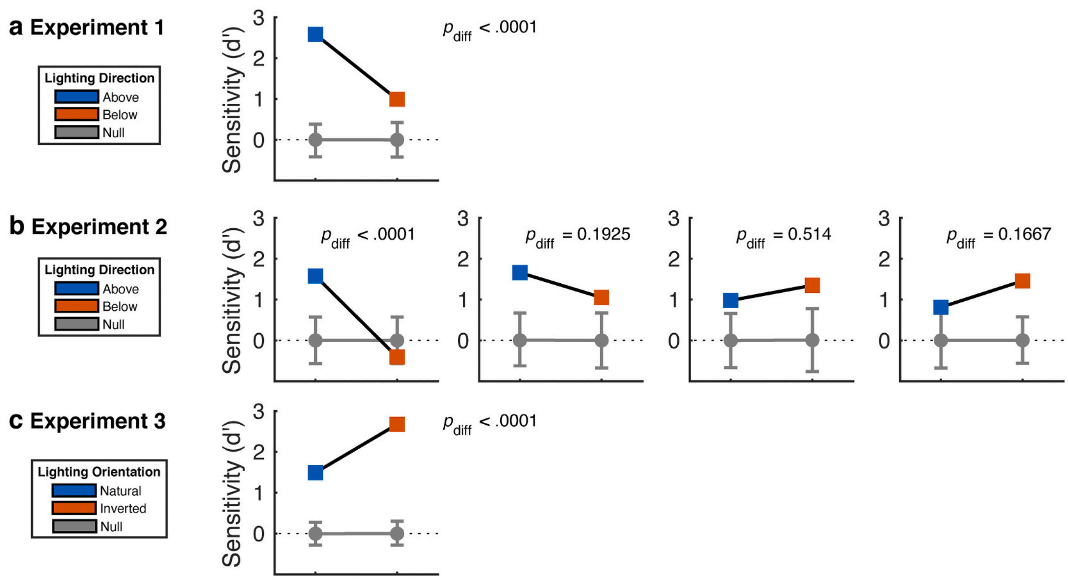

Fig. 10. Image classifier performance as a measure of the information content in two-tone images. **(a)** Classifier performance for images used in Experiment 1 with upright orientation and natural contrast polarity. The grey error bars indicate mean performance and 95% confidence intervals for a null distribution of classifier performance. p_{diff} values relate to the difference between classifier performance when trained and tested on images lit from above versus images lit from below. **(b)** Classifier performance for images used in Experiment 2 with upright orientation and threshold level (from left to right) 12.5%, 25%, 50%, 75%. **(c)** Classifier performance for images used in Experiment 3 with upright orientation.

direction interacts with the morphology of internal face features to produce an informative and familiar visual pattern. When an upright human face is lit from above, shadows tend to occur below the brows and nose, for example, producing a distinct pattern of contrast to that which occurs when faces are lit more strongly from below, illustrated in Fig. 8. There is evidence that a rudimentary, 2D template of contrast differences may play a key role in face detection, from studies of newborn looking behaviour (Farroni et al., 2005), illusory face detection (Paras & Webster, 2013; Smith, Gosselin, & Schyns, 2012), primate electrophysiology (Ohayon et al., 2012) and machine vision (Sinha, 2002). The current results demonstrate how basic sensory patterns of this nature that are useful for face detection are generated (in part) by the interaction between face shape and vertical asymmetries in lighting that commonly occur in natural environments.

5.1. The advantage for light from above in face detection

In real-world environments, light typically arrives from all directions at once, but often more strongly from above the horizon (Dror et al., 2004). There are several reasons why an advantage for detecting faces lit from above might arise in visual processing. First, there is evidence that the visual system has a default expectation for overhead lighting that contributes to our perception of object shape, particularly when the illumination conditions are ambiguous (Morgenstern et al., 2011; Ramachandran, 1988). In two-tone images, knowing the position of the light source may help to disambiguate object shape (Moore & Cavanagh, 1998) – similarly, an assumption for overhead lighting may aid the recovery of object shape from images that are consistent with overhead lighting, while potentially hindering the recovery of object shape from images that are inconsistent with overhead lighting. Consistent with this, Brodski et al. (2015) report that faces are more readily detected in two-tone images that are consistent with light arriving from above the observer, primarily when the face itself is oriented upside-down. Such an advantage may be a general feature of object recognition rather than face-processing per se, discussed further below in the section *Generalizability of lighting-dependent performance*.

A second reason why a light-from-above advantage might arise in face detection relates to the notion that the visual system employs a broadly-tuned template of visual features that are common to different faces, matching this template against incoming visual signals to detect the presence of a face in our environment. An advantage for light from above may reflect that this template is better-tuned to the appearance of the (upright) human face that we commonly see under overhead lighting (e.g., shadows under the brows; Farroni et al., 2005; Johnson, 2005). The results of the present study provide evidence to this effect. In particular, across all three experiments, we find that manipulating the lighting direction *relative to the face* produces considerable differences in face-detection performance, independent of the lighting direction relative to the observer. We are able to dissociate these by examining the effects of spatial inversion of the images – we find that faces lit from the direction of the brow are more commonly identified as faces than those lit from the direction of the chin, even when these images are presented to the observer upside down, meaning that the brow-lit faces are consistent with light arriving from *below the observer* (Fig. 2). Hence, our results demonstrate the role of overhead illumination in interacting with the unique 3D shape of the upright human face to *generate a familiar or informative sensory pattern* (e.g., the pattern of contrast produced by shadows falling under the brows and nose), which is often recognizable even when seen in a less typical spatial orientation. In contrast, this feature of the results is not well-explained by the notion of prior expectations about lighting direction (relative to the observer) influencing the ability to recover object shape.

Why did the current results differ from those reported by Brodski and colleagues? This previous work reports a slight increase in accuracy for detecting upright faces in two-tone images consistent with lighting from above rather than below, with high performance in both conditions (~94% vs. 92% accuracy). The strong differences in performance between lighting conditions observed in the current study, in contrast, might have emerged here if our tasks were better able to avoid ceiling effects. More difficult to explain is the discrepancy in findings for the interaction between spatial inversion and lighting direction. It is worth noting, however, some key methodological differences between studies that affect the specific visual features under study: (i) using 3D rendering, we isolated shading and shadows produced under different lighting conditions, such that the two-tone images used in the current study did not contain features attributable to surface reflectance (e.g., the irises, eyebrows, and lips reflecting light at a different intensity to the surrounding skin; Fig. 1), (ii) we cropped the face and non-face images such that there was no recognizable external contour provided by the ears, hair, and neck, instead focusing on the role of internal features in generating the pattern of a face, and (iii) we measured sensitivity in discriminating faces from non-face objects under matched illumination, such that the non-face images used in the current study contained some comparable visual cues to the face images (e.g., the broad asymmetry in intensity that occurs across an ellipsoidal object under directional lighting, Fig. 4). Hence, it is plausible that participants were using different visual features to detect faces in the current study compared to the experiment of Brodski and colleagues, which may have contributed to the different effects of lighting direction observed. In particular, the current results suggest that when considering the pattern of contrast produced by shading and shadowing across the internal features of the face, the lighting direction relative to the face is more significant in generating a recognizable sensory pattern than the lighting direction relative to the observer.

What are the visual features that drive face detection in two-tone images? As noted in the previous paragraph, our stimulus design allows us to attribute face detection performance in the current experiments to the broad patterns of contrast produced by shading and shadows that occur across the internal features of the face under directional lighting. The particular contrast patterns associated with different lighting conditions are illustrated in Figs. 8 and 9a – these highlight that dark patches tend to occur under the brows, nose, and lips in image sets associated with better performance. Moreover, we found that an increase in performance for faces lit from below the chin, when the two-tone images were thresholded at a high intensity level, was associated with a comparable pattern of edges to that which appeared to facilitate performance in faces lit from above the brow (in particular, edges occurring at the level of the eyes, nostrils, and lips; Fig. 9b). In ecological conditions, there are additional cues likely to contribute to face detection that were not available in our two-tone images, including, for example, patterns of contrast produced by the non-uniform reflectance of the facial surface (e.g., the characteristic appearance of the eyes produced by the dark iris and lighter sclera) and the external contours of the human head. Hence, the contrast patterns that we use as stimuli in the current study are not fully representative of those that occur across faces in ecological viewing conditions – instead, the present results isolate a contribution of shading and shadow patterns to the appearance of the face that appears familiar-enough to observers to enable face detection, at least when consistent with light arriving more strongly from above the brow.

Another factor that may contribute to differences in face detection and object recognition across lighting environments is the amount of low-level information available in the image (e.g., the quantity of edges generated across the surface of the object under different lighting conditions). We quantified low-level information in our two-tone images in terms of how well an image classifier could learn to distinguish faces from non-face objects. Interestingly, this analysis indicated that the amount of useful information for making this discrimination varies depending on the vertical lighting direction, speaking to how a (vertically) asymmetric object like the human face can generate either more- or less-consistent sensory patterns under different lighting. Classifier performance partly aligned, in a qualitative manner, with human performance. A notable exception was for images rendered in complex lighting environments; here, humans were better at detecting faces under naturalistic lighting despite the classifier finding *more* diagnostic information in images produced in an inverted lighting environment. This suggests that the human advantage for naturalistic lighting is not explained by the amount of low-level information available in the images alone, but might also reflect familiarity with the patterns of shading typically encountered on faces.

5.2. Dependence on intensity-threshold level

We generally found that patterns of contrast produced when light arrives from above the brow facilitated face detection. The one exception was for two-tone images that were produced by thresholding faces at a high intensity level (the 75% threshold condition in Experiment 2), which isolates the pattern formed by the brightest regions of the original image. In this condition, human observers performed better in detecting faces in two-tone images that were consistent with lighting from below the chin rather than lighting from above the brow, both when the images were presented upright and upside down. Images that are thresholded at a high intensity level tend to capture parts of the face angled most directly towards the light source – this is due to the nature of shading, whereby the intensity of light reflected towards an observer from a given point of an object's surface depends on the orientation of the surface normal relative to the angle of incident light. This can be seen in Fig. 6A – the brightest regions of the image are distributed differently across the face when lit from above versus below, and potentially capture more relevant detail in the latter. One factor here is the spatial asymmetry of faces: finer details occur towards the bottom of the face (mouth, nostrils) compared to the top of the face (forehead), which impacts on the image features captured by the highest-intensity regions of the image when the face is lit from above versus below. Interestingly, there was a notable similarity between the image features that appeared to provide an advantage in face detection in the light-from-above and light-from-below conditions – in both lighting conditions, better behavioural performance across threshold levels was associated with edges occurring at the level of the eyes, nostrils, and lips (Fig. 9b). This is consistent with there being a simple pattern of visual features supporting face detection, one that commonly occurs in the broad contrast pattern produced by shading and shadows when an upright face is lit from above (Experiment 1–3) but which can also be produced from faces lit from below under certain conditions (Experiment 2, 75% threshold condition). Overall, the contrast patterns produced by overhead lighting appear to facilitate face detection, but the interaction between form and lighting can provide circumstances where light arriving from below gives more face information.

5.3. Generalizability of lighting-dependent performance

Are the current results specific to faces or do they reflect a more general characteristic of object recognition? A general advantage for recognizing objects in visual patterns that are consistent with light from above may occur if the visual system has a default expectation for light from above that is employed when attempting to infer the shape of an object from the retinal image (discussed, for instance, in Morgenstern et al., 2011). This could be implemented prior to object categorization and face detection. However, as discussed above, the interaction between lighting direction and spatial inversion of the face that we observe suggests that the advantage for light from above in our data (when viewing upright faces) is related primarily to the particular visual pattern produced when a face is lit from the direction of the brow, rather than a general advantage for shape recognition that occurs when the image is consistent with light from above the observer. Nevertheless, if the task were instead to detect a different category of object, it may still be reasonable to predict better performance for images consistent with light from above, if those are the lighting conditions that we commonly see the object type under in the real world. However, this may depend on whether we tend to see that object type with a specific orientation relative to the sun/sky (like faces and cars) and whether the object has sufficient asymmetry in its 3D structure to produce a notably different pattern of shadows and shading when lit from different angles. On the latter point, the features of the human face are asymmetrical along the vertical dimension, and this is partly why quite different patterns of shadows result from faces lit from above versus below – for example, deeper shadows are produced around the eyes when the face is lit from above because the brows tend to protrude over the eyes to a greater extent than the lower orbital area protrudes underneath the eyes. So, whether a given object type is better-recognized by the pattern of shadows that occurs specifically when it is lit from above will likely depend on the particular 3D shape of the object. Our results are evidence that the particular shape of the upright human face leads to differences in the basic contrast pattern produced when lit from above versus below that are significant enough to impact strongly on human performance in recognizing that pattern as a face.

This dependence of shading and shadowing on the specific 3D structure of the object also means that there are likely to be cross-species differences in how lighting direction contributes to the appearance of the face. Much research on the neural basis of face processing has been performed in non-human primates, such as macaque monkeys, including that which examines the tuning of cells in face-selective regions of the visual cortex to coarse contrast features (e.g., Ohayon et al., 2012). The tuning of such cells may derive in part from the pattern of shadows and shading that tend to occur across the face under naturalistic lighting (e.g., shadows below the brows when an upright face is lit more strongly from above). Non-human primates share many qualitative features of face morphology with humans, often including a brow that protrudes above the eyes, but also exhibit notable differences, such as a flatter nose and rounded muzzle. Hence, the pattern of shading and shadows produced across the facial surface when lit from above may be comparable to humans in some respects but not others, and this may be one factor that contributes to both similarities and potential differences in the visual features that drive face detection across species. This relates to a broader question regarding how the mechanics of face processing differ across primate species, discussed in the context of face recognition in Rossion and Taubert (2019).

In the experiments reported in the current paper, we assessed the tendency for participants to perceive an image as containing a face or not, while varying the lighting conditions used to generate the image. While the stimuli were presented very briefly, participants were not restricted in the time they had to respond following stimulus presentation, nor did we analyze response times or speed-accuracy tradeoffs. In principle, the speed of face detection may also differ across lighting conditions – for example, faster response times might occur for more familiar face-like patterns compared to less familiar patterns that are still identifiable as a face (e.g., Brodski et al., 2015). Previous work has also assessed aspects of face detection by measuring saccadic reaction times to peripheral targets and detection-times for targets presented under interocular suppression (e.g., Stein et al., 2011; Tomalski et al., 2009).

5.4. Conclusion

Our experience of the social world is built upon specialized sensory pathways that extract information about other people's appearance and behaviour (Pitcher & Ungerleider, 2021). To engage our 'social brain' we must detect people around us, requiring neural mechanisms for discriminating faces from other visual patterns, which appear to partly exploit coarse intensity differences that commonly occur across the human face (Tsao & Livingstone, 2008). Tuning to the essential features of a face might be built from simpler orientation- and polarity-sensitive filters in early- and mid-level visual cortex (e.g., V1 and V4; 't Hart et al., 2011; Ohayon et al., 2012) and/or a subcortical pathway (Johnson, 2005). Here we find that face detection in broad patterns of contrast depends strongly on the vertical lighting direction relative to the face, specifically in how this produces a recognizable pattern of shading and shadows. These results speak to the importance of shading and shadows in generating the sensory cues that drive face detection, and the adaptation of face detection mechanisms to the statistics of real-world illumination.

CRediT authorship contribution statement

Colin J. Palmer: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – original draft, Funding acquisition. **Erin Goddard:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Colin W.G. Clifford:** Conceptualization, Methodology, Writing – review & editing, Resources.

Declaration of Competing Interest

None.

Acknowledgements

This work was supported by an Australian Research Council Discovery Early Career Researcher Award (DE190100459) to CP.

Appendix B. Supplementary data

Supplementary Figs. S1–S5 are available in a supplementary file. The raw data that supports this article are also uploaded as supplementary material. Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105172.

References

't Hart, B. M., Abresch, T. G., & Einhauser, W. (2011). Faces in places: Humans and machines make similar face detection errors. *PLoS One, 6*(10). https://doi.org/10.1371/journal.pone.0025373. e25373.
Adini, Y., Moses, Y., & Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 721–732.
Brodski, A., Paasch, G. F., Helbling, S., & Wibral, M. (2015). The faces of predictive coding. *The Journal of Neuroscience, 35*(24), 8997–9006. https://doi.org/10.1523/JNEUROSCI.1529-14.2015
Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (Ed.), *Representations of vision: Trends and tacit assumptions in vision research* (pp. 295–304).
Cavanagh, P., & Leclerc, Y. G. (1989). Shape from shadows. *Journal of Experimental Psychology. Human Perception and Performance, 15*(1), 3–27. https://doi.org/10.1037//0096-1523.15.1.3
Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world illumination. *Journal of Vision, 4*(9), 821–837. https://doi.org/10.1167/4.9.11
Enns, J. T., & Shore, D. I. (1997). Separate influences of orientation and lighting in the inverted-face effect. *Perception & Psychophysics, 59*(1), 23–31. https://doi.org/10.3758/bf03206844
Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America, 102*(47), 17245–17250. https://doi.org/10.1073/pnas.0502205102
Hasan, M. K., Ahsan, M. S., Abdullah-Al-Mamun, Newaz, S. H. S., & Lee, G. M. (2021). Human face detection techniques: A comprehensive review and future research directions. *Electronics, 10*(2354), 1–46.
Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews. Neuroscience, 6*(10), 766–774. https://doi.org/10.1038/nrn1766
Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception, 21*(3), 365–375. https://doi.org/10.1068/p210365
Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*(1), B1–11. https://doi.org/10.1016/s0010-0277(98)00035-3
Kuroki, D. (2020). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods, 53*(301–310).
Lange, K., Kuhn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of Web Servers Supporting Online Studies. *PLoS One, 10*(6). https://doi.org/10.1371/journal.pone.0130834. e0130834.
de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y
Leo, I., & Simion, F. (2009). Newborns' Mooney-face perception. *Infancy, 14*(6), 641–653. https://doi.org/10.1080/15250000903264047
Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology, 11*(4), 219–226. https://doi.org/10.1037/h0083717
Moore, C., & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition, 67*(1–2), 45–71. https://doi.org/10.1016/s0010-0277(98)00014-6

Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences of the United States of America, 108*(30), 12551–12553. https://doi.org/10.1073/pnas.1100794108

Ohayon, S., Freiwald, W. A., & Tsao, D. Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron, 74*(3), 567–581. https://doi.org/10.1016/j.neuron.2012.03.024

Omer, Y., Sapir, R., Hatuka, Y., & Yovel, G. (2019). What is a face? Critical features for face detection. *Perception, 48*(5), 437–446. https://doi.org/10.1177/0301006619838734

Otsu, N. (1979). A threshold selection method from gray-level histograms. In *, 1. IEEE transactions on systems, man, and cybernetics*.

Paras, C. L., & Webster, M. A. (2013). Stimulus requirements for face perception: An analysis based on "totem poles". *Frontiers in Psychology, 4*, 18. https://doi.org/10.3389/fpsyg.2013.00018

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences, 25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006

Python Core Team.. (2017). *Python: A dynamic, open source programming language*. Retrieved from https://www.python.org/psf/.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*(6152), 163–166. https://doi.org/10.1038/331163a0

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025. https://doi.org/10.1038/14819

Rossion, B., & Taubert, J. (2019). What can we learn about human individual face recognition from experimental studies in monkeys? *Vision Research, 157*, 142–158. https://doi.org/10.1016/j.visres.2018.03.012

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 411–426. https://doi.org/10.1109/TPAMI.2007.56

Sinha, P. (2002). Qualitative representations for recognition. In H. H. Bülthoff, C. Wallraven, S. W. Lee, & T. A. Poggio (Eds.)*, 2525. Biologically Motivated Computer Vision. BMCV 2002. Lecture Notes in Computer Science.* (pp. 249–262). Berlin, Heidelberg: Springer.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology, 22*(3), 191–196. https://doi.org/10.1016/j.cub.2011.11.061

Stein, T., Peelen, M. V., & Sterzer, P. (2011). Adults' awareness of faces follows newborns' looking preferences. *PLoS One, 6*(12). https://doi.org/10.1371/journal.pone.0029361. e29361.

Tomalski, P., Csibra, G., & Johnson, M. H. (2009). Rapid orienting toward face-like stimuli with gaze-relevant contrast information. *Perception, 38*(4), 569–578. https://doi.org/10.1068/p6137

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *, 1. Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition* (pp. 511–518).