

Cognition and Behavior

Modulations of Depth Responses in the Human Brain by Object Context: Does Biological Relevance Matter?

Idy W. Y. Chou,¹ Hiroshi Ban,^{2,3} and Dorita H. F. Chang^{1,4}<https://doi.org/10.1523/ENEURO.0039-21.2021>

¹Department of Psychology, The University of Hong Kong, Hong Kong, China, ²Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, Osaka, 565-0871, Japan, ³Graduate School of Frontier Biosciences, Osaka University, Osaka, 565-0871, Japan, and ⁴State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China

Abstract

Depth sensitivity has been shown to be modulated by object context (plausibility). It is possible that it is behavioral relevance rather than object plausibility per se which drives this effect. Here, we manipulated the biological relevance of objects (face or a non-face) and tested whether object relevance affects behavioral sensitivity and neural responses to depth-position. In a first experiment, we presented human observers with disparity-defined faces and non-faces, and observers were asked to judge the depth position of the target under signal-noise and clear (fine) task conditions. In the second experiment, we concurrently measured behavioral and fMRI responses to depth. We found that behavioral performance varied across stimulus conditions such that they were significantly worse for the upright face than the inverted face and the random shape in the signal-to-noise (SNR) task, but worse for the random shape than the upright face in the feature task. Pattern analysis of fMRI responses revealed that activity of fusiform face area (FFA) was distinctly different during depth judgments of the upright face versus the other two stimuli, with its responses (and to a stronger extent, those of V3) appearing functionally-relevant to behavioral performance. We speculate that FFA is not only involved in object analysis, but exerts considerable influence on stereoscopic mechanisms as early as in V3 based on a broader appreciation of the stimulus' behavioral relevance.

Key words: depth perception; fMRI; object recognition

Significance Statement

We asked how disparity sensitivity is modulated by object (biological) relevance using behavioral and neuroimaging paradigms. We show that behavioral sensitivity to depth-position changes in biological (face) versus non-biological (random surface) contexts, and that these changes are task-dependent. Imaging results highlight a potentially key role of the fusiform region for governing the modulation of stereo encoding by object relevance. These findings highlight powerful interactions between object recognition mechanisms and stereoencoding, such that the utility of disparity information may be up/down weighed depending on the biological relevance of the object.

Introduction

Binocular disparity, referring to the difference in the left- and right-eye's image that is a consequence of the two

eyes' separation on the head, is crucial to our everyday ability to recognize and interact with objects (Burke, 2005; Dal Mullo et al., 2011). Neurophysiological work has

Received January 28, 2021; accepted June 10, 2021; First published June 17, 2021.

The authors declare no competing financial interests.

Author contributions: I.W.Y.C., H.B., and D.H.F.C. performed research; D.H.F.C. conceptualized and designed the research; I.W.Y.C., H.B., and D.H.F.C. analyzed data; I.W.Y.C., H.B., and D.H.F.C. wrote the paper.

indicated a wide network of regions in the primate brain, starting from V1, and extending dorsally to V3 and the MT complex, as well as ventrally to V4 and IT, that respond selectively to binocular disparity (Hubel and Wiesel, 1970; Poggio and Fischer, 1977; Poggio and Talbot, 1981; Poggio et al., 1988; Roy et al., 1992; Cumming and Parker, 1997, 1999; DeAngelis et al., 1998; Uka et al., 2000; Hinkle and Connor, 2001). Many of these same areas are responsible for more complex representations rendered by stereoscopic cues. For instance, the caudal intraparietal (CIP) region along dorsal cortex, and V4 and along ventral cortex, respond to 3D slant (Janssen et al., 1999, 2000; Hegdé and Van Essen, 2005; Sakata et al., 2005; Durand et al., 2007).

More recent studies suggest that the particular region engaged during the perception of depth from disparity depends on the task employed. In the primate, the MT complex appears to respond during judgments of disparity signals from noise (DeAngelis et al., 1998; Uka and DeAngelis, 2003, 2004, 2006). On the other hand, V4 and IT seem to be central to the discrimination of fine, but clear depth differences (Uka et al., 2005; Shiozaki et al., 2012). This apparent segregation of the roles of dorsal and ventral regions in different depth discrimination tasks has been observed in humans as well (Chang et al., 2014).

Task-based modulation of stereo responses observed in the cortex suggests that the disparity-processing circuitry is sensitive to context. Recent work has shown that in addition to task requirements, disparity mechanisms can be sensitive to an entirely different form of context, object identity/class. In particular, depth sensitivity is surprisingly modulated by object plausibility (Wong et al., 2020). Specifically, during depth position-judgment tasks, performance was worse for physically plausible versus implausible objects. This effect is particularly striking as both classes of stimuli were equivalently complex, and perhaps more importantly, the plausibility of the object was task-irrelevant. Wong et al. (2020) demonstrated further that the physical plausibility of the object appeared to modulate disparity-related multivariate response patterns in intermediate and higher dorsal (V3A, V3B, V7) and ventral (lateral occipital; LO) regions. These findings highlight the importance of object-level information in processing stereoscopic signals.

Here, we sought to better understand the nature of the dependence of depth responses on object context by asking whether the biological relevance of objects can affect behavioral sensitivity and neural responses to

stereoscopic depth. We manipulated the relevance of objects by presenting stereoscopic face stimuli (at upright and inverted orientations) and non-face surfaces matched in depth-power (i.e., amplitude). Indeed, object (biological) relevance has been shown to play a role in the perception of many visual attributes. For instance, the perceived color of food items is biased toward the familiar color of the object stored in long-term memory (Hansen et al., 2006; Olkkonen et al., 2008). Moreover, the race category of a face can affect the perceived lightness of the face (Levin and Banaji, 2006; Chang et al., 2018). The human visual system's apparent sensitivity to faces is of course now well known to be served by dedicated regions for processing these stimuli, including the fusiform face area (FFA; Kanwisher et al., 1997; McCarthy et al., 1997), the superior temporal sulcus (Puce et al., 1995, 1996), and the occipital face area (Gauthier et al., 2000; Pitcher et al., 2007). Interestingly, the illusory percepts of faces are sufficient to elicit activity in these same regions (Liu et al., 2014).

The apparent importance of faces to our visual system suggests that it is an appropriate stimulus context to be used to explore the effect of biological relevance on disparity processing. In two experiments, we investigated how behavioral sensitivity (experiment 1) and fMRI responses (experiment 2) to depth are modulated by biological relevance. We created random-dot stereograms (RDSs) depicting faces (upright and inverted) and depth-matched random shapes. We elected to include an upside-down variant of the face stimulus as it is well established that faces are more poorly perceived when they are presented inverted (Yin, 1969). Turning the face upside-down thus, presents an interesting stimulus condition that allows us to preserve the structural integrity of an upright face, but is yet reduced in terms of its biological relevance.

Materials and Methods

Experiment 1 (behavior)

In experiment 1, we measured depth thresholds in two tasks: a signal-in-noise [signal-to-noise ratio (SNR)] task of judging the depth position (near/far) of a central target relative to the surrounding plane when the SNR varied between 0% and 100%, and a feature task that required discriminating which of two consecutively presented clear objects was nearer. We deemed it important to index performance according to both tasks as they have been shown to engage distinct regions in cortex (Chang et al., 2014; Patten and Welchman, 2015).

Participants

A total of 60 observers were recruited for this first experiment: 30 participants (aged 19–34 years; mean 24.9 years; 10 males), were tested on the SNR task and a separate group of thirty participants (aged 18–34 years; mean 22.0 years; 12 males) participated in the feature task. All participants were screened for stereo deficits (Titmus test), had normal or corrected-to-normal vision as screened with the Snellen linear acuity chart, and provided written informed consent in line with ethical

This work was supported by Early Career Scheme (27612119) and General Research Fund (17612920) grants (Research Grants Council, Hong Kong) to D. C.; by the Foreign Researcher Invitation Program (National Institute of Information and Communications Technology, Japan) to D.C.; and by Japan Society for the Promotion of Science Grants KAKENHI (17H04790 and 21H00968) and ERATO JPMJER1801 to H.B.

Correspondence should be addressed to Dorita H. F. Chang at changd@hku.hk.

<https://doi.org/10.1523/ENEURO.0039-21.2021>

Copyright © 2021 Chou et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

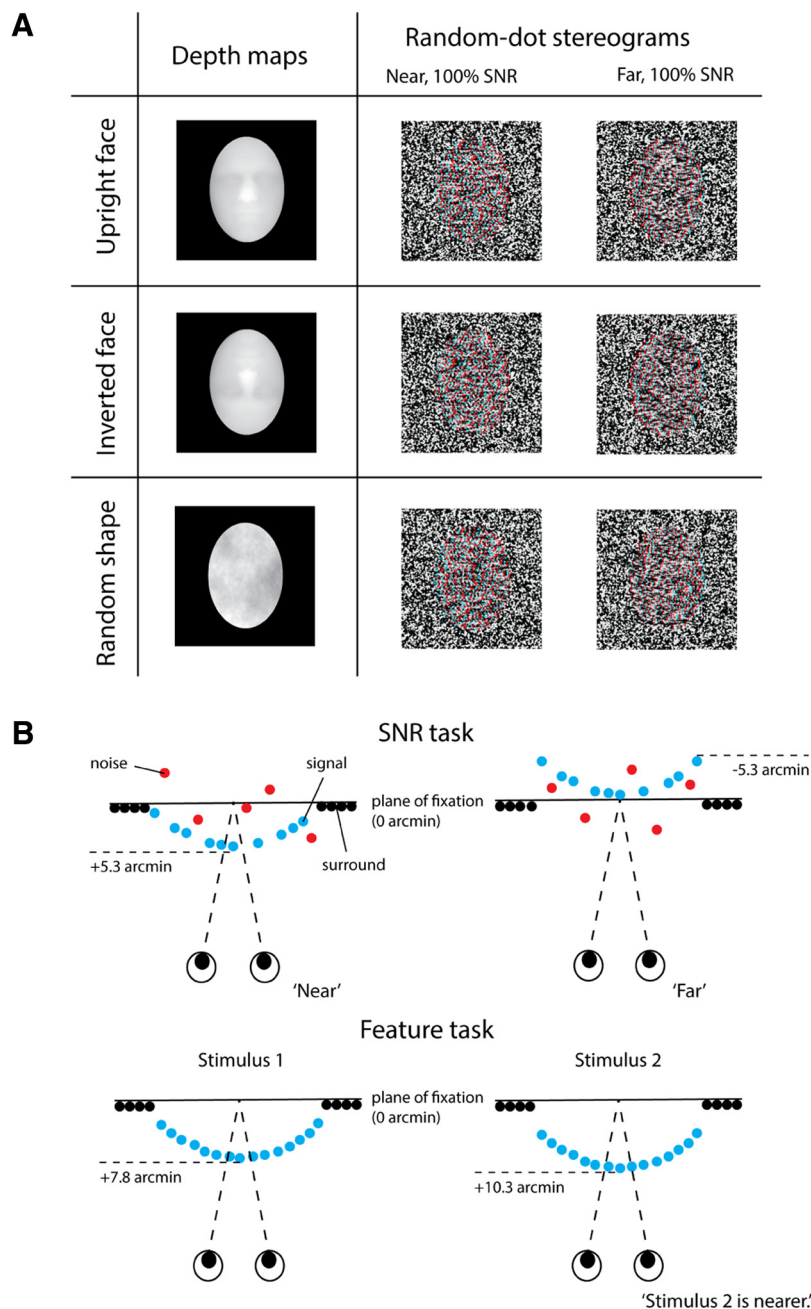


Figure 1. A, Sample depth maps and random-dot stereograms (RDS). Top, The depth map of the upright face was presented in an oval region that cropped the ears and the neck. Center, An inverted face was created by rotating the upright face by 180° in the picture plane. Bottom, A depth-power matched random shape was generated by randomizing the phases of the Fourier transform of the upright face while preserving the amplitudes. Anaglyphic examples of the RDS in “near” and “far” conditions, at 100% SNR, are also shown for illustration purposes only. **B**, In the signal-noise (SNR) task (top), SNR varied between 0% and 100%. In the feature task (bottom), two stimuli were presented consecutively on each trial. Disparity difference varied within 150 arcs.

approval by the institutional Human Research Ethics Committee (HREC).

Apparatus

Stimuli were generated using MATLAB (MathWorks) with extensions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and presented through a mirror stereoscope, in which two eyes viewed the left and right halves of a 24-inch monitor (1920 × 1080 resolution;

60-Hz refresh rate) through four front-surface mirrors. Viewing distance was 65 cm, and a chinrest was used to limit head movements and maintain viewing distance.

Stimuli

Three kinds of stimuli, including the upright face, the inverted face, and the depth-power-matched random shape, were rendered as RDSs (Fig. 1A). 3D laser-scanned head models were firstly obtained from the Face

Database of the Max Planck Institute for Biological Cybernetics (Troje and Bühlhoff, 1996; Blanz and Vetter, 1999). All heads were in frontal view with hair removed. Four face identities, including two male and two female head models, were randomly selected and rendered as depth maps such that intensity coded for depth. The size and shape of all faces were standardized by applying an oval-shape cropping mask that eliminated the ears and the neck. The inverted faces were generated by rotating the upright faces by 180° in the picture plane. The depth-power-matched random shapes were created by taking the depth map of each upright face, applying a Fourier transform, and randomizing its phases while preserving the amplitudes. The resulting images thus carry no structural coherence but maintain identical depth information as the upright and inverted variants. A total of four random shapes were created, each corresponding to one face identity.

The depth maps were finally rendered as RDSs, where the amount of pixel-wise disparity was indexed by intensity. Each RDS had a maximum disparity of ± 5.3 arcmin and a size of $8.68^\circ \times 8.68^\circ$, within which the target subtended 6.45° vertically and 4.55° horizontally. Each target had an average disparity of ± 4.27 arcmin. Dots were randomly assigned as black or white, with a size of 0.05° and a density of 20 dots/deg². The RDS was presented against a black background and surrounded by a grid of gray and white squares (each 0.5° in size) that served to provide an unambiguous background reference and stabilize vergence position.

Tasks

SNR task. On each trial, participants were required to judge whether a central target was “near” or “far” as compared with the surrounding region (the “surround”). Disparity of the surround was fixed at zero (i.e., at the plane of fixation). For the target, disparity was fixed at ± 5.3 arcmin at the maximum point across the target’s surface. Task difficulty was manipulated by varying the SNR, altering the proportion of dots that defined the target (“signal”) relative to noise dots that were assigned a random disparity within the range of ± 5.3 arcmin (Fig. 1B). The initial test value was 80% SNR. On each trial subsequently, the SNR of the stimulus was adjusted according to the QUEST staircase procedure, estimating thresholds at the 82% correct level (Watson and Pelli, 1983).

Feature task. On each trial, participants were required to judge which of two consecutively presented objects was “nearer”. One of the objects was designated the “reference” and had a fixed disparity of 7.8 arcmin. Task difficulty was manipulated by varying the disparity difference between the target and the reference within a range of 150 arcs (Fig. 1B). Both the target and reference had crossed disparity and were presented at 100% SNR. The initial test value of the target was set at 120 arcs (relative to the reference). As for the SNR task, on each trial, disparity difference was adjusted according to the QUEST staircase procedure yielding threshold estimates at the 82% accuracy level (Watson and Pelli, 1983).

Before each task, participants were provided with oral instructions, viewed examples of the stimuli, and completed twelve practice trials with audio feedback to get familiarized with the tasks. Each task consisted of six test runs [three conditions (upright face, inverted face, random-surface) \times two stimulus identities] presented in random order. A test run contained one type of object with two interleaved staircases of stimulus gender (i.e., the gender of the head models from which the stimuli were derived). Each staircase had 64 trials that comprised of four practice trials and 60 test trials. The test value of the first test trial was determined by the threshold obtained in the practice trials.

In both tasks, each trial began with a nonius-type fixation box (0.25° in size) that consisted of dichoptically presented nonius lines (each eye perceives only two of the four lines), which lasted for 500 ms. In the SNR task, each stimulus was presented for 300 ms. In the feature task, each stimulus was presented for 300 ms with an interstimulus interval of 500 ms. In both tasks, stimulus presentation was followed by a response period that ended after a response was made or 3000 ms, whichever came first. Trials without a response after the maximum response period elapsed were deemed incorrect. Participants were required to respond using the arrow keys on a computer keyboard.

Experiment 2 (fMRI)

In experiment 2, we measured fMRI responses when subjects were performing the SNR task, and examined both univariate and multivariate responses to depth under the different stimulus conditions. In this second experiment, we selected to test the SNR task as it elicited the more robust behavioral effects both here, and in previous work (Wong et al., 2020).

Participants

Twenty-two participants (aged 21–34; mean 23.2 years; 17 males) who did not participate in the behavior experiments were recruited for the fMRI experiment. All participants had normal or corrected-to-normal vision and provided written informed consent in line with ethical approval by the institutional HREC.

Apparatus

Left and right images were back-projected using two projectors (WUX4000, Canon), each equipped with a polarizing filter (resolution: 1280×1024 pixels; 60-Hz refreshing rate) and placed at 96 cm from the back of the fMRI bore. Participants viewed the stimuli with the corresponding polarized glasses through a 45° tilted mirror mounted in front of the head.

Stimuli

Stimuli were the same as those used in experiment 1 except for the following differences. First, for each condition (upright face/inverted face/random shape), only two stimulus exemplars were used, one per stimulus gender, resulting in a total of six unique images used in this experiment. This change was made to minimize the noise of fMRI signals across stimulus conditions that could result

with large stimulus variance. Second, the target subtended 5.23° vertically and 3.70° horizontally to accommodate a smaller field of view in the bore setup. Third, the maximum disparity on the target surface was 5.4 arcmin.

fMRI acquisition

Imaging data were acquired at the Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Osaka, Japan, using a 3-T Siemens Trio MR scanner with a half of the 32-channel, phase-array (whole) head coil that covered the occipital lobe. Head movements were limited by a foam padding inside the coil. For both localizer and experimental runs, blood oxygenation level-dependent (BOLD) signals were measured with a multiband echoplanar imaging (EPI) sequence [voxel size = $2 \times 2 \times 2 \text{ mm}^3$, echo time (TE) = 30 ms, repetition time (TR) = 2000 ms, field of view = 192×192 , flip angle = 75° , 78 axial slices (slightly oblique along the AC-PC line), multiband factor = 3; 205 volumes where the first 5 were discarded to eliminate the effects of start-up transients] provided by the University of Minnesota (under a C2P contract). Additionally, high-resolution T1-weighted images were collected for each participant (voxel size = 1 mm^3 , TE = 2.48 ms, TR = 1900 ms, field of view = 256×256 , flip angle = 9° , 208 slices).

Region of interest (ROI) definition

For each participant, we defined ROIs V1, V2, V3, V4, V3A, and V7 using standard phase-encoded retinotopic mapping procedures that mapped polar angle with a checkerboard wedge stimulus that was rotating periodically (clockwise or anti-clockwise; <https://github.com/hiroshiban/Retinotopy>; Sereno et al., 1995). Area V3B/KO (kinetic occipital) was defined using the same map, as the region falling anterior to V3A and inferior to V7 (DuPont et al., 1997). Using a single-run functional localizer, we identified hMT+ (human motion complex) as a set of contiguous voxels that showed significantly stronger responses ($p < 0.01$) to an array of moving dots that receded or expanded coherently than to an array of static dots (Huk et al., 2002). Note that because of time constraints, five participants did not complete the hMT+ localizer scan. For these participants, we defined hMT+ as fixed 5-mm spherical ROIs centered on Talairach coordinates of $[-51, -72, 0]$ and $[51, -69, 3]$ in the left and right hemispheres, respectively (Orban et al., 2003). LO was identified as the region in the lateral occipito-temporal cortex that was significantly more responsive ($p < 0.01$) to intact than scrambled images of objects and shapes (Kourtzi and Kanwisher, 2001). Bilateral FFAs were defined as 5-mm spherical ROIs for all participants, with Talairach coordinates of $[-37, -42, -16]$ in the left hemisphere and $[39, -40, -16]$ in the right hemisphere (Grill-Spector et al., 2004). Note that FFA was not localized in individual subjects as we saw no reason to extract responses from solely face-stimulus-specific voxels in this region.

Design and procedures

Before scanning, each participant firstly completed three runs of the SNR task inside the MRI bore (with the

scanner idle). Thresholds estimated from these in-bore behavior-only runs were used to determine individually-tailored stimulus test values for subsequent fMRI acquisition runs. This was done to standardize task difficulty across conditions and participants, thus reducing the noise of fMRI signals in our data. From these behavior-only runs, we obtained subject-specific test values for the fMRI runs as follows: for each staircase (stimulus condition), we took the average of the test values in last 30 trials and defined a range of ± 1 SD from this mean value. Subsequently, on each fMRI trial (of the corresponding condition), the SNR value was randomly sampled within this range. For any instances where a single staircase could not yield reliable estimates [because of: (1) a lack of asymptotic convergence, and/or (2) 95% confidence intervals reaching floor] stimuli were sampled at 80% SNR in the corresponding fMRI trials. This implicated a total of 19 (of 132 total staircases) run in bore. fMRI runs were arranged in a block design. Each run comprised seven block types, corresponding to six stimulus conditions [3 (type) \times 2 (gender)] and fixation. Each block lasted for 16 s. A stimulus block comprised 8 trials with equal number of “near” and “far” trials. On each trial, a stimulus was presented for 500 ms and was followed by a fixed-duration response period of 1500 ms. A background grid of randomly gray and white squares was present throughout the entire run.

A fixation block contained a nonius-type fixation (see above for details). In each run, each stimulus block type was repeated twice, with condition order randomized, and interleaved with fixation blocks. Thus, there were a total of 12 stimulus blocks and 13 fixation blocks (each run began and ended with a fixation block). Each run consisted of 96 trials and lasted 6.4 min. The entire scanning session lasted around 90 min.

fMRI data analysis

Data were processed using BrainVoyager QX (Goebel et al., 2006; Formisano et al., 2006). For each participant, T1-weighted anatomic scans were transformed into Talairach space (Talairach and Tournoux, 1988). Inflated and flattened cortical surfaces were reconstructed for both hemispheres. Functional data were preprocessed with slice time correction, 3D rigid-body motion correction, linear trend removal and high-pass filtering (frequency space filter; three cycles per run) without spatial smoothing.

We performed two analyses that examined both univariate [general linear model (GLM)] responses and pattern-level responses using multivoxel pattern analysis (MVPA). The GLM consisted of three regressors representing the three stimulus conditions (upright face, inverted face and random shape) and six regressors of motion parameters (three translation parameters and three rotation parameters) and a constant term. Stimulus-condition regressors were defined by modeling stimulation periods separately for each stimulus condition by a boxcar model convolved with a (two- γ) hemodynamic response function (HRF; Glover, 1999). The averaged time course signal across all voxels of each ROI was then modelled as a linear combination of the regressors. The regressor coefficients of

different stimulus conditions and ROIs were then contrasted.

MVPA classification analysis was done with a linear support vector machine implemented in libSVM tool (Chang and Lin, 2011) with a recursive feature elimination (RFE) technique (De Martino et al., 2008). The RFE is a multivariate feature selection algorithm that eliminates uninformative voxels. The time course signals of all voxels were converted into z scores. Also, to account for the typical delay of the hemodynamic response, signals were shifted by 4 s (= 2 TRs; Serences, 2004). For each ROI, the SVM was trained to classify the patterned responses between the upright face and the random shape, between the inverted face and the random shape, and between the upright face and the inverted face.

A leave-one-run-out cross-validation procedure was used for the MVPA. Specifically, in each cross-validation, functional data of one run were used as test dataset and the rest of the data were used as training dataset for computing SVM weights. In each RFE step, the SVM classifier was trained with 80% of the training dataset, which were resampled 20 times without partition of blocks. Therefore, each voxel was assigned with 20 SVM weights. All voxels were ranked according to the average weights in that particular RFE step, and the five voxels that had the lowest average weights were eliminated. The remaining data were used to decode the test pattern, and the prediction accuracy for that particular voxel pattern (in the particular RFE step) was obtained. This procedure was repeated until the voxel count fell below 250, after which it would proceed to the next cross-validation. As such, the final pattern size for all ROIs here was 250 voxels. Note that the analysis was in fact repeated at all possible fine voxel counts between 50 and 800 voxels (stepping in 50 voxel increments), but we ultimately decided to report results from the 250 pattern size as this corresponded to the smallest pattern size at which accuracies reached asymptotic levels. For each ROI and classification, the final prediction accuracy was tested against permutation baseline level (0.53), as determined by running 1000 SVMs with shuffled labels. In addition to an ROI-based MVPA, we conducted a searchlight analysis in which the pattern discriminability between stimulus conditions was tested by moving a 6-mm (radius) spherical ROI across the cortex.

We explored the relationship between MVPA classification accuracies and behavioral thresholds. For each set of comparisons [(1) upright face vs random shape, (2) inverted face vs random shape, and (3) upright face vs inverted face], we extracted its SVM accuracy and calculated a behavioral index, corresponding to the difference in behavioral thresholds between the two conditions in the SNR task. We then computed the (Pearson's) correlation coefficient (r) between the two metrics for each ROI. We also performed a further analysis concatenating the inverted face and random shape conditions, forming a "non-upright-face" condition. The corresponding SVM was run, and accuracies subsequently correlated with a behavioral index similarly computed by averaging thresholds from the two non-upright-face conditions.

Results

Behavior

Behavioral performance for the SNR and feature tasks, indexed in terms of signal-noise (SNR task) or disparity difference thresholds (feature task) are presented in Figure 2. Sample staircases from one subject are shown in Figure 2C. Thresholds for each task were analyzed using a one-way repeated-measures ANOVA comparing thresholds across stimulus conditions (Table 1). The analysis for the SNR task (Fig. 2A) indicated a significant main effect of stimulus condition, $F_{(2,58)} = 5.65$, $p = 0.006$, $\eta_p^2 = 0.163$. *Post hoc* Bonferroni-corrected t tests revealed that thresholds were higher (poorer) for the upright face than the inverted face, $t_{(29)} = 3.00$, $p = 0.018$, $d = 0.547$ and for the upright face than the random shape, $t_{(29)} = 2.89$, $p = 0.021$, $d = 0.528$. However, there was no significant difference between the thresholds for the inverted face and that for the random shape $t_{(29)} = 0.337$, $p > 0.999$, $d = 0.067$.

The analysis for feature task thresholds (Fig. 2B) revealed a significant main effect of stimulus condition, $F_{(2,58)} = 4.26$, $p = 0.019$, $\eta_p^2 = 0.128$. *Post hoc* Bonferroni-corrected t tests indicated that thresholds were significantly lower (better) for the upright face than the random shape $t_{(29)} = 2.79$, $p = 0.027$, $d = 0.510$, but there was no significant difference between the thresholds for the inverted face and the upright face, $t_{(29)} = 0.431$, $p > 0.999$, $d = 0.079$, and that between the inverted face and the random shape, $t_{(29)} = 2.04$, $p = 0.153$, $d = 0.372$.

fMRI

In-bore behavior (SNR task)

We verified firstly that behavioral effects for the task chosen for the MRI could be replicated by the new group of participants while inside the bore. A one-way repeated-measures ANOVA indicated a significant effect of stimulus condition, $F_{(2,36)} = 3.45$, $p = 0.043$, $\eta_p^2 = 0.161$. *Post hoc* tests indicated that, consistent with findings from experiment 1, thresholds for the upright face were significantly higher than those for the random shape: $t_{(18)} = 2.33$, $p = 0.032$, $d = 0.448$, but there were no significant differences between thresholds for the upright face and the inverted face ($t_{(18)} = 1.74$, $p = 0.098$, $d = 0.244$), and between the inverted face and the random shape ($t_{(18)} = 1.05$, $p = 0.309$, $d = 0.169$).

Turning to the threshold-sampled stimuli of the functional runs, we then assessed whether the difference in disparity distribution between stimuli at the "near" and "far" stimulus positions (relative to the fixation plane) might have an influence on response accuracies by means of a 2 (depth-position: near/far) \times 3 (condition) ANOVA. The analyses indicated no significant main effect of depth-position, $F_{(1,21)} = 0.053$, $p = 0.820$, $\eta_p^2 = 0.003$, nor interactions involving depth-position [mean accuracy (SD) for Near_{uprface} = 0.81 (0.13); Far_{uprface} = 0.79 (0.18); Near_{invface} = 0.80 (0.17); Far_{invface} = 0.81 (0.15); Near_{ran} = 0.78 (0.13); Far_{ran} = 0.82 (0.13)].

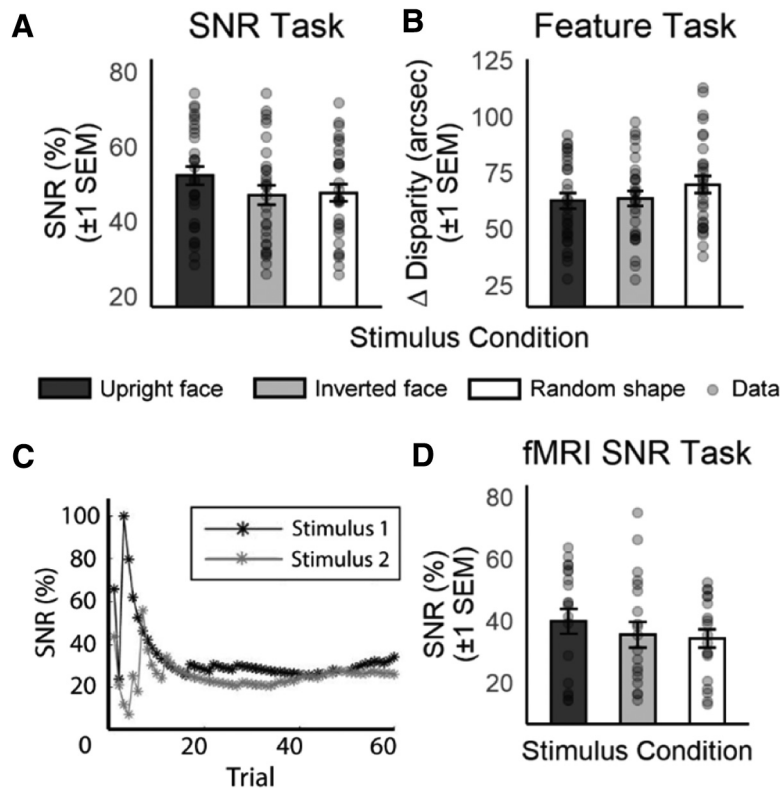


Figure 2. A, B, Depth discrimination thresholds in the signal-noise (SNR) and feature tasks. In the SNR task, thresholds were higher (i.e., poorer) for the upright face than the inverted face and the random shape. In the feature task, thresholds for the upright face were significantly lower (i.e., better) than those for the other two stimulus conditions, whereas thresholds for the inverted face did not differ significantly from those of the other two stimulus conditions. **C,** Sample staircases from a randomly-sampled observer, showing threshold estimates across trials within one experimental run, for the upright face condition. **D,** Behavioral thresholds obtained in-bore (SNR task). Thresholds were higher for the upright face than the random shape. Individual subject thresholds are overlaid in **A–C**. Error bars represent ± 1 SEM.

GLM

We firstly examined univariate fMRI signals across stimulus conditions. For each hemisphere, we extracted the GLM β -weights (percent signal change) corresponding to the three conditions from the ten ROIs (Fig. 3). These were then entered into a 2 (hemisphere) \times 3 (type) \times 10 (ROI) repeated-measures ANOVA (Table 1) that revealed significant main effects of hemisphere ($F_{(1,21)}=5.28$, $p=0.032$, $\eta_p^2=0.201$) and ROI ($F_{(9,189)}=7.60$, $p<0.001$, $\eta_p^2=0.266$). β -Weights did not differ across conditions ($F_{(2,42)}=1.87$, $p=0.167$, $\eta_p^2=0.082$), and there were no significant interactions: hemisphere \times condition: $F_{(2,42)}=0.609$, $p=0.549$, $\eta_p^2=0.028$; hemisphere \times ROI: $F_{(9,189)}=0.656$, $p=0.748$, $\eta_p^2=0.030$; ROI \times condition: $F_{(18,378)}=0.610$, $p=0.892$, $\eta_p^2=0.028$; hemisphere \times condition \times ROI: $F_{(18,378)}=0.137$, $p=0.141$, $\eta_p^2=0.036$. A comparison of the mean β -weights indicated that signals were generally stronger in the left hemisphere ($M=0.145$, $SE=0.083$) than the right hemisphere ($M=0.073$, $SE=0.079$).

Note that while the GLM results indicated stronger responses in the left hemisphere as compared with the right hemisphere, amplitudes differences were homogenous across ROIs and stimulus conditions. For this reason, and, as subsequent multivariate analyses remove the univariate amplitude, we elected to

concatenate data from the two hemispheres for subsequent analyses.

MVPA

Multivariate responses to different stimulus conditions were examined by contrasting the multivoxel fMRI signals between the upright face and the random shape, and those for the inverted face and the random shape (see Materials and Methods). MVPA classification accuracies are presented in Figure 4A. Classification accuracies were compared with baseline (0.53) while holding familywise error rate at 0.05 (qFDR < 0.05 ; Table 1).

SVM accuracies for the classification between the upright face and the random shape were significantly above permutation-baseline in V2, V3, V7, V4, LO, and FFA, while those between the inverted face and the random shape were above-baseline in V1, V2, V3A, V7, V4, and LO. For the comparison between the upright face and inverted face, SVM accuracies were above-baseline in V3A, V3B, V4, and FFA. In other words, brain regions that had above-baseline classification performance in the contrasts involving the random-shape (which may reflect simply the classification of differences in lower-order features) included V2, V7, V4, and LO. Regions that showed above-baseline discriminability for the contrast between

Table 1: Statistical table

Analysis	Figure	Data structure	Type of test	Variables	Observed power
Experiment 1, SNR task	2A	Assumed normal	One-way repeated-measures ANOVA	Independent: stimulus condition Dependent: depth threshold	0.843
Experiment 1, feature task	2B	Assumed normal	One-way repeated-measures ANOVA	Independent: stimulus condition Dependent: depth threshold	0.722
Experiment 2, behavior	2D	Assumed normal	One-way repeated-measures ANOVA	Independent: stimulus condition Dependent: depth threshold	0.610
Experiment 2, GLM	3	Assumed normal	Three-way repeated-measures ANOVA	Independent: hemisphere Stimulus condition ROI Dependent: GLM β -weight	0.580 0.372 >0.999
Experiment 2, MVPA	4	Assumed normal	One-sample <i>t</i> test	SVM accuracy against baseline Upright face vs random shape classification	
				V1	0.690
				V2	>0.999
				V3	0.972
				V3A	0.744
				V3B	0.422
				V7	0.919
				hMT+	0.703
				V4	0.880
				LO	0.905
				FFA	0.762
				Inverted face vs random shape classification	
				V1	0.959
				V2	0.997
				V3	0.639
				V3A	0.999
				V3B	0.479
				V7	0.998
				hMT+	0.175
				V4	0.999
				LO	0.764
				FFA	0.396
				Upright face vs inverted face classification	
				V1	0.437
				V2	0.418
				V3	0.064
				V3A	0.999
				V3B	0.875
				V7	0.504
				hMT+	0.088
				V4	0.903
				LO	0.540
				FFA	0.838
Experiment 2, MVPA-behavior correlation	4	Assumed normal	Pearson correlation	Variable 1: Threshold difference Variable 2: SVM accuracy Upright face vs random shape	
				V1	0.050
				V2	0.378
				V3	0.288
				V3A	0.071
				V3B	0.055
				V7	0.074
				hMT+	0.051
				V4	0.690
				LO	0.258
				FFA	0.075
				Inverted face vs random shape	
				V1	0.312

(Continued)

Table 1: Continued

Analysis	Figure	Data structure	Type of test	Variables	Observed power
				V2	0.087
				V3	0.087
				V3A	0.579
				V3B	0.117
				V7	0.342
				hMT+	0.053
				V4	0.053
				LO	0.075
				FFA	0.072
				Upright face vs inverted face	
				V1	0.533
				V2	0.449
				V3	0.984
				V3A	0.058
				V3B	0.353
				V7	0.250
				hMT+	0.202
				V4	0.488
				LO	0.193
				FFA	0.544
				Upright face vs non-upright-face	
				V1	0.057
				V2	0.129
				V3	0.170
				V3A	0.060
				V3B	0.110
				V7	0.065
				hMT+	0.061
				V4	0.121
				LO	0.064
				FFA	0.562

the upright face and the random shape only were V3 and FFA, while those above-baseline for the contrast between the inverted face and the random shape only were V1 and V3A. Only V3B uniquely distinguished between responses of the upright versus inverted faces. Finally, only FFA exhibited above-baseline discrimination accuracies that were common and unique to both SVMs involving the upright face stimulus (i.e., above-baseline for both the upright face vs random, and upright face vs inverted face comparisons, but not the inverted vs random comparison).

Lastly, the searchlight analysis reflected that our choice of ROIs well covered the regions showing above-baseline pattern discriminability. Sample searchlight maps for the pattern discriminability (t value) between the upright face and random shape conditions are presented in Extended Data Figure 4-1. Note that these searchlight maps were not further used to emphasize our main findings together with the other results to avoid circularity problems (Kriegeskorte et al., 2009).

Brain-behavior relationships

Finally, to clarify the relevance of these brain areas for the effects observed behaviorally, we further examined the relationship between behavioral thresholds and MVPA accuracies within each ROI (Fig. 4) in two sets of analyses. First, for each possible comparison of the three stimulus conditions [(1) upright face vs random shape, (2)

inverted face vs random shape, (3) upright face vs inverted face], we derived a behavioral index computed as the difference of SNR thresholds between the two conditions. We then calculated the (Pearson's) correlation coefficient (r) between this behavioral index and the SVM accuracies from the classification of the corresponding stimulus conditions for each ROI across all participants. We found that for the "upright face versus random shape" comparison, SVM accuracies in V4 correlated positively to the behavioral index ($r = 0.513$, $p = 0.025$); for the "inverted face versus random shape" comparison, SVM accuracies in V3A correlated negatively to behavioral index ($r = -0.465$, $p = 0.045$); and finally, for the "upright face versus inverted face" comparison, SVM accuracies in V3 correlated positively to the behavioral index ($r = 0.707$, $p < 0.001$). Amongst these results, only the last involving V3 withstood Bonferroni correction for multiple comparisons.

Second, we performed an additional MVPA-behavior correlational analysis with data of the inverted face and random shape conditions combined to form a "non-upright-face" condition. Indeed, previous work has suggested that upright faces represent a class of objects that is distinct from non-face objects, as well as inverted faces (Yin, 1969; Valentine, 1988; Farah et al., 1995; Kanwisher et al., 1997; Kanwisher and Yovel, 2006). Behavioral indices (upright face minus non-upright-face conditions) were then correlated with results from a separate MVPA that

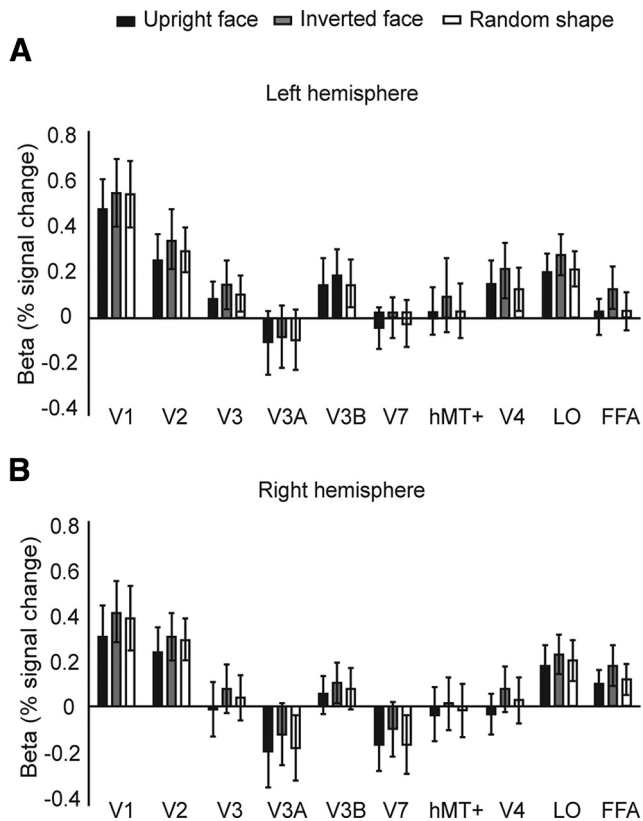


Figure 3. GLM β -weights (% signal change) of the three stimulus conditions in the (A) left and (B) right hemispheres. Univariate signals were globally higher in the left hemisphere than the right hemisphere. Signals were generally higher in V1, V2, and LO. Error bars represent ± 1 SEM ($N = 22$).

compared classification accuracies for discriminating between the upright face and non-upright-face (i.e., inverted face + random shape) conditions. We found a significant correlation between behavior and SVM accuracies in FFA only ($r = 0.481$, $p = 0.041$; Extended Data Fig. 4-2), although the correlation did not survive further statistical correction.

Discussion

We tested for differences in behavioral sensitivity and fMRI responses to stereoscopic depth, in face and non-face contexts. In experiment 1, we compared behavioral depth discrimination performances under three conditions with varying biological relevance: upright faces, inverted faces, and phase-randomized surfaces. We showed that the behavioral sensitivity to depth differed across conditions, but curiously, also across tasks. In the SNR task, thresholds were higher for the upright face than the inverted face and the random shape; that is, SNR depth sensitivity was poorer in a biologically relevant context. In contrast, in the feature task, depth thresholds were higher for the random shape than the upright face; that is, fine thresholds were instead, better for biologically relevant stimuli. In experiment 2, we selected the SNR task, as it elicited more robust

differences in responses across stimulus contexts, both in experiment 1 of the present study and in previous work (Wong et al., 2020), and examined fMRI responses to the stimuli concurrently with behavior. Analysis of the univariate responses showed rather homogenous activation across stimulus conditions. However, results from the MVPA were more revealing. Early and intermediate retinotopic areas (V1, V2, V3, V3A, V3B, V4) and extrastriate dorsal (V7) and ventral (LO, FFA) regions showed discriminable response patterns across contexts. Notably, of these regions, responses of FFA in particular appear to be classifying information beyond simple lower-order features (i.e., are uniquely and commonly discriminative for contrasts that involved the upright face). Moreover, multivariate responses of V3, V4 and FFA are predictive (albeit weakly) of behavioral responses. We consider the behavioral and imaging findings in turn.

Modulation of behavioral sensitivity to depth by object relevance

We firstly consider our behavioral findings. Our data indicate that depth sensitivity differs between face and non-face contexts and, importantly, this difference is task-dependent. Our results showed that a face context, in particular in the upright orientation, facilitates feature (disparity) discrimination but worsens the segmentation of disparity signal from noise. The reversed trends observed between the tasks, at first glance, may seem paradoxical. However, as it has been proposed that performance in tasks similar to those employed here can be thought to involve two different mechanisms, namely noise filtering and feature readout (Doshier and Lu, 2005), we propose that the differences between the SNR task and feature task can be explained within this framework. The level of external noise is high in the SNR task but absent in the feature task, implying that noise filtering is the dominating visual process during the SNR task but not the feature task. The two tasks also require different levels of precision: the SNR task requires a lower level of precision (near/far depth discrimination), whereas the feature task requires a higher level of precision (discrimination of fine depth difference). This implies a heavy engagement in feature readout for the feature task, but only a modest amount of feature extraction could suffice depth judgments in the SNR task.

Why then, is noise filtering better in non-face objects, and feature readout better in face contexts? First, that the detection of a target from noise is better for novel, unfamiliar objects has been well documented in the literature. For example, when participants are presented with an array of four objects and then asked to recall the location of an item in the array, accuracy is significantly higher when the test item is novel than when it is familiar (Johnston et al., 1990). This “novel popout” effect, referring to the fact that novel objects are easier to detect when noise is present, may relate to the higher saliency of novel (vs familiar) objects that captures attention and hence facilitates target detection. In our SNR task, the

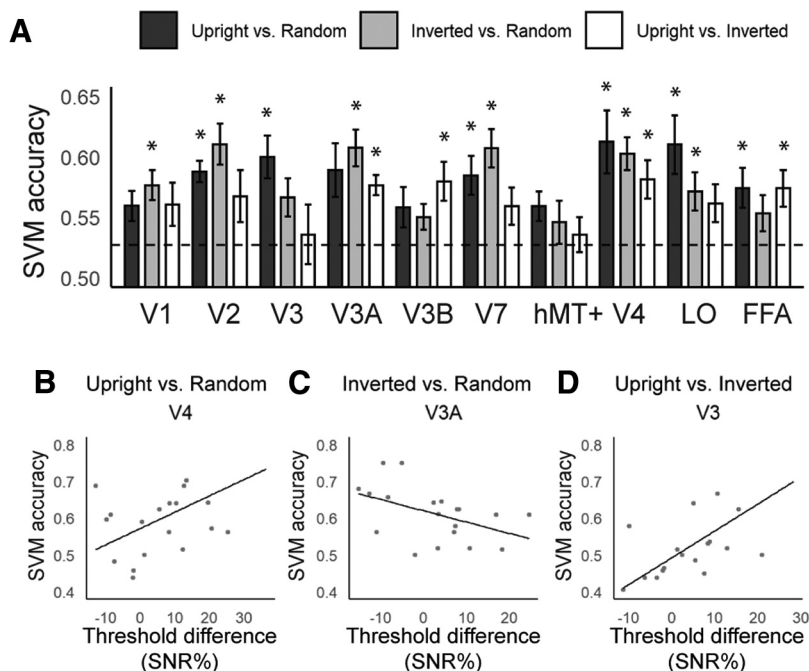


Figure 4. **A**, SVM accuracies for all ROIs in the three exhaustive classifications comparing all stimulus conditions. Classification accuracies (Bonferroni-corrected) that are significantly above-baseline (0.527; dashed line) are indicated by asterisks. Error bars represent ± 1 SEM ($N=22$). **B–D**, Brain-behavior correlations in V4 for upright-face versus random-shape (**B**); V3A for inverted-face versus random-shape (**C**); and V3 for upright-face versus inverted-face (**D**). The horizontal axis represents a behavioral index computed as the threshold difference between conditions. Additional results from the whole-brain classification (searchlight) and from an additional brain-behavior correlation following the concatenation of the inverted and non-face conditions (FFA) are presented in Extended Data Figures 4-1, 4-2, respectively.

“novel popout” effect can be manifest as an improvement in the quality of signal extraction for the random shape (compared with the upright face), thus reducing thresholds. Furthermore, it has been shown that the detection of faces appears to be especially susceptible to noise compared with other non-face daily objects (Chen et al., 2015). In particular, the presence of noise impairs feature discrimination to a larger extent for faces than for cars. In the context of this previous work then, our data for the SNR task appear to be in alignment.

Second, we found that feature readout is better for the upright face than for the random shape. In everyday functioning, the readout of fine features is crucial to face individuation and the recognition of facial expressions. Indeed, that a higher level of precision is required for the perception of faces than other objects has been shown empirically. Participants inclined to name faces at the level of its identity (i.e., the name of the face), but dogs at the level of its species or subspecies (e.g., “golden retriever” or simply “dog”), indicating that the processing of faces, compared with other objects, is more specific (Tanaka, 2001). Other work has shown that human participants are strikingly sensitive to fine differences in faces (Lehky, 2000). It is thus very likely that fine feature readout is more relevant to everyday face perception tasks. By contrast, precise feature discrimination is rarely required for non-face objects. Here, it is important to note that any behavioral enhancements of feature readout in face

contexts shall not necessarily manifest in the results of the SNR task as only a limited amount of feature readout was needed in this task.

Finally, we observed a clear difference of thresholds between the upright and inverted faces in the SNR task. This is in line with the well-documented face inversion effect, in which the inversion of face affects perception, both in terms of recognition performance (Yin, 1969; Valentine, 1988; Farah et al., 1995) and neural responses (Kanwisher et al., 1998) of face-selective regions. Yet, behavioral thresholds did not differ substantially between the upright face and the inverted face in the feature task. Interestingly, these findings seem to suggest that the face inversion effect is more prominent in high noise, compared with the discrimination of fine feature differences in low visual noise.

Importantly, the opposite behavioral trends observed in our SNR and feature tasks distinguishes these findings from those observed by Wong et al. (2020), who showed that the physical plausibility of objects affects depth sensitivity similarly for SNR and feature discrimination tasks. Particularly, depth sensitivity was poorer for physically plausible than implausible objects. At first glance, our data are at odds with the findings of Wong et al. (2020). However, we view that the findings generated by both studies are in fact complementary to each other. The data from Wong et al. (2020) suggest a reweighing of stereo processes based on the physical plausibility of the object. They suggested that during the perception of familiar

(physically plausible) objects, stereo signals are down-weighted while object-level processing is up-weighted, thus optimizing the interpretation and interaction with these familiar, naturally relevant objects. While [Wong et al. \(2020\)](#) showed a general modulation of disparity processes by the object's context (physical plausibility), we argue that our data, on the other hand, suggest a more specific contextual effect on stereo processing, that depends on the biological relevance of the object. Here, all three kinds of stimulus used in our study, namely the upright face, the inverted face and the random shape, are physically plausible. If the stereo re-weighting mechanism as suggested by [Wong et al. \(2020\)](#) only cares about whether the stimulus is physically plausible or not, one should expect no sensitivity differences to be observed across our stimuli. On the other hand, if the re-weighting mechanism is based more broadly on familiarity, then, one should expect a down-weighting of stereo sensitivity for face, especially the upright face, compared with the inverted face and the random shapes. Our data in the SNR task are in line with this prediction. However, there is something more to unravel for the feature task, for which we observed a somewhat reversed behavioral trend. Here, we suggest that the everyday face perception tasks, such as face identification, recognition of facial expressions, etc., require the processing of precise details in the stimulus. Therefore, during the perception of faces, it would be likely that a pre-existing circuitry that favors detailed visual processing would be triggered, which in turn facilitates the readout of features. Critically, such facilitation in feature readout should be robust and able to counteract or even out-weigh the stereo down-weighting effect that results from the more general process as suggested by [Wong et al. \(2020\)](#).

fMRI responses during depth judgments under varying object contexts

We next turn to our fMRI data. The GLM revealed rather homogenous univariate activation across stimulus conditions during depth judgments. Intriguingly, there was an overall higher amplitude of responses in the left versus right hemisphere. We speculate that this asymmetry has nothing to do with stereoscopic vision per se, as indeed, existing work has indicated an absence of hemispheric effects ([Akay and Celebi, 2009](#)) in depth perception. We fathom instead that our results might have been caused by the strong imbalance of genders in our recruited participants, with many more males than females in our fMRI component. To this end, there is evidence of globally stronger hemispheric asymmetry in males than in females ([Hirnstein et al., 2019](#)).

Our MVPA results were more revealing. We reasoned that we may be able to detect functionally-relevant responses to face-object modulations of stereo sensitivity, by teasing out any regions that can uniquely and commonly discriminate response patterns for the two stimulus contrasts involving the upright face stimulus (i.e., upright face vs random shape; upright face vs inverted face), but not for the comparison involving the inverted face and random shape. This allows us to avoid, as far as possible,

picking up regions that might be responding to lower-order stimulus differences. We found that of our ROIs, only FFA showed unique above-baseline SVM accuracy for those classifications that involve the upright face, suggesting a potentially key role for this region for explaining the strong modulations of stereo-sensitivity by biological relevance that was observed behaviorally.

V3B, which was uniquely discriminative of upright face versus inverted face responses here, has been previously implicated in the discrimination of correlated and anticorrelated RDS ([Preston et al., 2008](#)). Both V3 ([Backus et al., 2001](#); [Anzai et al., 2011](#)) and the inferotemporal cortex ([Janssen et al., 2000, 2003](#)) more broadly, have also been well demonstrated to be stereosensitive. Of these two regions, V3 showed here particular relevance to behavior (i.e., the upright vs inverted face, brain-behavior correlation). Here, larger discriminability of the two stimuli exhibited by V3 translated to larger discrepancies in the behavioral sensitivities to the two stimuli. While modulations in early visual cortex, long believed to care only about local stimulus features appear perplexing, this finding appears to well fit previous work showing modulation of responses in V1 to ambiguous stereograms by changing perceived depth via luminance changes of the surround ([Rideaux and Welchman, 2019](#)).

Responses of retinotopic V1, V2, V3A, V7 found here likely reflect decoding of lower order structural differences between the objects during depth judgments as these regions showed above-baseline SVM accuracy in both classifications involving the random shape (upright face vs random shape and inverted face vs random shape). Further, above-baseline classification accuracies in the ventral regions (V4, LO) found in both comparisons likely reflect the decoding of the object's identity and/or lower-order stereo information: V4 has been shown to respond to absolute depth ([Shiozaki et al., 2012](#)) and structured objects, regardless of the exact object identity ([Desimone et al., 1985](#); [Merigan and Pham, 1998](#)). LO is well demonstrated to respond more strongly to intact than scrambled objects ([Malach et al., 1995](#)), and shows similar responses to familiar and novel shapes ([Margalit et al., 2016](#)). These findings are in line with our results where pattern discriminability was similar across classifications, although the upright and inverted faces differ in the level of familiarity.

We showed that depth sensitivity is affected by the biological relevance of the objects, defined here in terms of whether the target object depicted a face or non-face, and that such modulations are task-dependent. Using fMRI, we observed modulations of multivariate fMRI responses during depth judgments, particularly in FFA, by object type. We speculate that FFA is not only involved in object analysis as discussed above, but exerts considerable influence on stereoscopic mechanisms as early as in V3 such that the utility of disparity information is up/down regulated depending on the relevance of the object (i.e., whether it is a face or non-face). The role of FFA in face perception is well documented ([Kanwisher et al., 1997, 1998](#)), but its up-stream and down-stream interactions with more rudimentary processes (here, stereo-encoding,

in particular carried by V3) is well worth continued exploration, perhaps through using more temporally-resolved techniques. Lastly, we fathom that our findings in the FFA likely reflect our choice of biological manipulations here (i.e., faces or non-faces). In this vein, it would be further interesting to explore object-stereo interactions under different indices that vary biological importance.

References

- Akay A, Celebi G (2009) Depth perception differences between hemispheres, genders, and eye dominance: a study on evoked potential. *Neural Regen Res* 4:947–950.
- Anzai A, Chowdhury SA, DeAngelis GC (2011) Coding of Stereoscopic Depth Information in Visual Areas V3 and V3A. *J Neurosci* 31:10270–10282.
- Backus BT, Fleet DJ, Parker AJ, Heeger DJ (2001) Human cortical activity correlates with stereoscopic depth perception. *J Neurophysiol* 86:2054–2068.
- Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp 187–194). Los Angeles, California USA, July.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Burke D (2005) Combining disparate views of objects: viewpoint costs are reduced by stereopsis. *Visual Cogn* 12:705–719.
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27.
- Chang DHF, Mevorach C, Kourtzi Z, Welchman AE (2014) Training transfers the limits on perception from parietal to ventral cortex. *Curr Biol* 24:2445–2450.
- Chang DHF, Cheang YY, So M (2018) Contextual effects in face lightness perception are not expertise-dependent. *Vision (Basel)* 2:23.
- Chen Y, McBain R, Norton D (2015) Specific vulnerability of face perception to noise: a similar effect in schizophrenia patients and healthy individuals. *Psychiatry Res* 225:619–624.
- Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389:280–283.
- Cumming BG, Parker AJ (1999) Binocular neurons in V1 of awake monkeys are selective for absolute, not relative. *J Neurosci* 19:5602–5618.
- Dal Mutto C, Zanuttigh P, Cortelazzo GM, Mattoccia S (2011). Scene segmentation assisted by stereo vision. *Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission* (pp 57–64). Hangzhou, China, May.
- DeAngelis GC, Cumming BG, Newsome WT (1998) Cortical area MT and the perception of stereoscopic depth. *Nature* 394:677–680.
- De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43:44–58.
- Desimone R, Schein SJ, Moran J, Ungerleider LG (1985) Contour, color and shape analysis beyond the striate cortex. *Vision Res* 25:441–452.
- Dosher BA, Lu ZL (2005) Perceptual learning in clear displays optimizes perceptual expertise: learning the limiting process. *Proc Natl Acad Sci USA* 102:5286–5290.
- Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, Marchal G, Mortelmans L, Orban GA (1997) The kinetic occipital region in human visual cortex. *Cereb Cortex* 7:283–292.
- Durand J, Nelissen K, Joly O, Wardak C, Todd JT, Norman JF, Janssen P, Vanduffel W, Orban GA (2007) Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron* 55:493–505.
- Farah MJ, Tanaka JW, Drain HM (1995) What causes the face inversion effect? *J Exp Psychol Hum Percept Perform* 21:628–634.
- Formisano E, Di Salle F, Goebel R (2006) Fundamentals of data analysis methods in fMRI. In: *Advanced image processing in magnetic resonance imaging* (Landini L, Positano V, and Santarelli MF, eds), pp 481–503. Boca Raton: CRC Press.
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform “face area” is part of a network that processes faces at the individual level. *J Cogn Neurosci* 12:495–504.
- Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9:416–429.
- Goebel R, Esposito F, Formisano E (2006) Analysis of FIAC data with BrainVoyager QX: from single-subject to cortically aligned group GLM analysis and self-organizing group ICA. *Hum Brain Mapp* 27:392–401.
- Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci* 7:555–562.
- Hansen T, Olkkonen M, Walter S, Gegenfurtner KR (2006) Memory modulates color appearance. *Nat Neurosci* 9:1367–1368.
- Hegd  J, Van Essen DC (2005) Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *J Neurophysiol* 94:2856–2866.
- Hinkle DA, Connor CE (2001) Disparity tuning in macaque area V4. *Neuroreport* 12:365–369.
- Hirstein M, Hugdahl K, Hausmann M (2019) Cognitive sex differences and hemispheric asymmetry: a critical review of 40 years of research. *Laterality* 24:204–252.
- Hubel DH, Wiesel TN (1970) Stereoscopic vision in macaque monkey. *Nature* 225:41–42.
- Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci* 22:7195–7205.
- Janssen P, Vogels R, Orban GA (1999) Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proc Natl Acad Sci USA* 96:8217–8222.
- Janssen P, Vogels R, Orban GA (2000) Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288:2054–2056.
- Janssen P, Vogels R, Liu Y, Orban GA (2003) At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron* 37:693–701.
- Johnston WA, Hawley KJ, Plewe SH, Elliot JMG, Dewitt MJ (1990) Attention capture by novel stimuli. *J Exp Psychol Gen* 119:397–411.
- Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci* 361:2109–2128.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kanwisher N, Tong F, Nakayama K (1998) The effect of face inversion on the human fusiform face area. *Cognition* 68:B1–B11.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by human lateral occipital complex. *Science* 293:1506–1509.
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540.
- Lehky SR (2000) Fine discrimination of faces can be performed rapidly. *J Cogn Neurosci* 12:848–855.
- Levin DT, Banaji MR (2006) Distortions in the perceived lightness of faces: the role of race categories. *J Exp Psychol Gen* 135:501–512.
- Liu J, Li J, Feng L, Li L, Tian J, Lee K (2014) Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. *Cortex* 53:60–77.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA* 92:8135–8139.

- Margalit E, Shah MP, Tjan BS, Biederman I, Keller B, Brenner R (2016) The lateral occipital complex shows no net response to object familiarity. *J Vis* 16:3.
- McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci* 9:605–610.
- Merigan WH, Pham HA (1998) V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. *Vis Neurosci* 15:359–367.
- Olkkonen M, Hansen T, Gegenfurtner KR (2008) Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *J Vis* 8:13.1–16.
- Orban GA, Fize D, Peuskens H, Denys K, Nelissen K, Sunaert S, Todd J, Vanduffel W (2003) Similarities and differences in motion processing between the human and macaque brain: evidence from fMRI. *Neuropsychologia* 41:1757–1768.
- Patten ML, Welchman AE (2015) FMRI activity in posterior parietal cortex relates to the perceptual use of binocular disparity for both signal-in-noise and feature difference tasks. *PLoS One* 10:e0140696.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Pitcher D, Walsh V, Yovel G, Duchaine B (2007) TMS evidence for the involvement of the right occipital face area in early face processing. *Curr Biol* 17:1568–1573.
- Poggio GF, Fischer B (1977) Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurosci* 40:1392–1405.
- Poggio GF, Talbot WH (1981) Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol* 315:469–492.
- Poggio GF, Gonzalez F, Krause F (1988) Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci* 8:4531–4550.
- Preston TJ, Li S, Kourtzi Z, Welchman AE (2008) Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J Neurosci* 28:11315–11327.
- Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol* 74:1192–1199.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 16:5205–5215.
- Rideaux R, Welchman AE (2019) Contextual effects on binocular matching are evident in primary visual cortex. *Vision Res* 159:76–85.
- Roy JP, Komatsu H, Wurtz RH (1992) Disparity sensitivity of neurons in monkey extrastriate area MST. *J Neurosci* 12:2478–2492.
- Sakata H, Tsutsui K, Taira M (2005) Toward an understanding of the neural processing for 3D shape perception. *Neuropsychologia* 43:151–161.
- Serences JT (2004) A comparison of methods for characterizing the event related BOLD timeseries in rapid fMRI. *Neuroimage* 21:1690–1700.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau J, Brady T, Rosen B, Tootell R (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893.
- Shiozaki HM, Tanabe S, Doi T, Fujita I (2012) Neural activity in cortical area V4 underlies fine disparity discrimination. *J Neurosci* 32:3830–3841.
- Talarach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. New York: Thieme Medical Publishers, Inc.
- Tanaka JW (2001) The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen* 130:534–543.
- Troje NF, Bühlhoff HH (1996) Face recognition under varying poses: the role of texture and shape. *Vision Res* 36:1761–1771.
- Uka T, DeAngelis GC (2003) Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J Neurosci* 23:3515–3530.
- Uka T, DeAngelis GC (2004) Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron* 42:297–310.
- Uka T, DeAngelis GC (2006) Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. *J Neurosci* 26:6791–6802.
- Uka T, Tanaka H, Yoshiyama K, Kato M, Fujita I (2000) Disparity selectivity of neurons in monkey inferior temporal cortex. *J Neurophysiol* 84:120–132.
- Uka T, Tanabe S, Watanabe M, Fujita I (2005) Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *J Neurosci* 25:10796–10802.
- Valentine T (1988) Upside-down faces: a review of the effect of inversion upon face recognition. *Br J Psychol* 79:471–491.
- Watson AB, Pelli DG (1983) Quest: a Bayesian adaptive psychometric method. *Percept Psychophys* 33:113–120.
- Wong NHL, Ban H, Chang DHF (2020) Human depth sensitivity is affected by object plausibility. *J Cogn Neurosci* 32:338–352.
- Yin RK (1969) Looking at upside-down faces. *J Exp Psychol* 81:141–145.