Differential processing of binocular and monocular gloss cues in human visual cortex

Hua-Chun Sun[1], Massimiliano Di Luca[1], Hiroshi Ban[2,3], Alexander Muryy[4], Roland W. Fleming[5], Andrew E. Welchman[6,*]

1. School of Psychology, University of Birmingham, UK
2. Center for Information and Neural Networks, National Institute of Information and Communications Technology, and Osaka University, Japan
3. Graduate School of Frontier Biosciences, Osaka University, Japan
4. School of Psychology, University of Southampton, UK
5. Department of Psychology, Justus-Liebig-Universität Gießen, Germany
6. Department of Psychology, University of Cambridge, UK

*Corresponding author.

E-mail address: aew69@cam.ac.uk (A. E. W.)

Abstract

The visual impression of an object's surface reflectance ('gloss') relies on a range of visual cues, both monocular and binocular. While previous imaging work has identified processing within ventral visual areas as important for monocular cues, little is known about cortical areas involved in processing binocular cues. Here we used human functional magnetic resonance imaging (fMRI) to test for brain areas selectively involved in the processing of binocular cues. We manipulated stereoscopic information to create four conditions that differed in their disparity structure and in the impression of surface gloss that they evoked. We performed multi-voxel pattern analysis to find areas whose fMRI responses allow classes of stimuli to be distinguished based on their depth structure vs. material appearance. We show that higher dorsal areas play a role in processing binocular gloss information in addition to known ventral areas involved in material processing, with ventral area LO responding to both object shape and surface material properties. Moreover, we tested for similarities between the representation of gloss from binocular cues and monocular cues. Specifically, we tested for transfer in the decoding performance of an algorithm trained on glossy vs. matte objects defined by either binocular or by monocular cues. We found transfer effects from monocular to binocular cues in V3B/KO, suggesting a shared representation of the two cues in this area. These results indicate the involvement of mid-to-high level visual circuitry in the estimation of surface material properties, with V3B/KO potentially playing a role in integrating monocular and binocular cues.

Keywords

Surface gloss; material perception; specularity; MVPA; fMRI; binocular cues

Introduction

Surface gloss provides important information about the characteristics of visual objects: for instance, shiny metal objects are usually recently manufactured and have better conductance than rusty metal, while fresh apples have glossier skin than rotten ones. However, estimating gloss poses a difficult challenge to the visual system: the viewer has to separate the surface properties of the object from information about the illumination and 3D shape of the object (Anderson, 2011). Here we sought to investigate the neural circuits that play a role in meeting this challenge to estimate gloss.

A number of investigators have studied the neural basis of gloss computations by manipulating the specular- and diffuse- surface reflectance properties of objects (Kentridge, Thomson, & Heywood, 2012; Nishio, Goda, & Komatsu, 2012; Nishio, Shimokawa, Goda, & Komatsu, 2014; Okazawa, Goda, & Komatsu, 2012; Sun, Ban, Di Luca, & Welchman, 2015; Wada, Sakano, & Ando, 2014). For instance, functional magnetic resonance imaging (fMRI) and single-cell recordings in the macaque brain have demonstrated that gloss information from reflections of the surrounding environment (i.e., specular reflections) is processed along ventral visual pathway from V1, V2, V3, V4 to superior temporal sulcus (STS) and inferior temporal (IT) cortex (Nishio et al., 2012; Okazawa et al., 2012). Similarly, human studies suggested that specular highlight cues to gloss are primarily processed in the ventral processing stream: V4, ventral–occipital (VO-1/VO-2) area, lateral occipital (LO) area, collateral sulcus (CoS) and posterior fusiform sulcus (pFs) (Sun et al., 2015; Wada et al., 2014). Further, these human studies suggested the involvement of dorsal visual area V3B/KO in gloss processing.

This previous work has involved participants looking at (stereoscopically) flat pictorial representations of glossy surfaces. This follows the tradition of psychophysical

studies that have identified a number of pictorial signals that could be used to identify surface reflectance properties (Anderson & Kim, 2009; Doerschner et al., 2011; Doerschner, Maloney, & Boyaci, 2010; Fleming, Dror, & Adelson, 2003; Gegenfurtner, Baumgartner, & Wiebel, 2013; Kim & Anderson, 2010; Kim, Marlow, & Anderson, 2011; Kim, Marlow, & Anderson, 2012; Landy, 2007; Marlow & Anderson, 2013; Marlow, Kim, & Anderson, 2011; Motoyoshi, Nishida, Sharan, & Adelson, 2007). For convenience, we will refer to these types of pictorial cues as 'monocular' in the sense that they allow a viewer to gain an impression of surface gloss based on a single view of the stimuli.

In addition to monocular gloss cues, it is clear that potentially important information about surface reflectance properties comes from binocular cues. In particular, viewing glossy surfaces binocularly typically results in the two eyes registering a different pattern of reflections, such that specular reflections are displaced away from the physical surface in depth (Blake & Bülthoff, 1990; Kerrigan & Adams, 2013; Wendt, Faul, & Mausfeld, 2008). Past psychophysical work has shown that these binocular signals can strongly modulate the impression of surface gloss (Blake & Bülthoff, 1990; Kerrigan & Adams, 2013; Muryy, Fleming, & Welchman, 2012; Obein, Knoblauch, & Viénot, 2004; Sakano & Ando, 2010; Wendt, Faul, Ekroll, & Mausfeld, 2010; Wendt et al., 2008). For instance, Blake & Bülthoff (1990) showed that simply changing the disparity of a highlight with respect to a physical surface could lead to a considerable change in participants' perceptual impression of surface gloss. Moreover, work characterizing the properties of binocular reflections has shown that the disparities evoked by such stimuli often differ substantially from the disparities evoked when viewing matte objects: disparity gradients are larger and there can be large vertical offsets between corresponding image features (Muryy, Fleming, & Welchman, 2014; Muryy, Welchman, Blake, & Fleming, 2013).

Here we sought to test for cortical areas engaged by monocular and binocular cues to gloss. The logic of our approach was to contrast stimuli that differed in binocular disparity structure or material appearance, and thereby localise fMRI responses to (i) disparity vs. (ii) perceived gloss. An ideal stimulus set would therefore contain (i) items that had the same material appearance but different disparity structures; and (ii) the same disparity but different material appearance, while in all cases keeping other image features identical. While this idea scenario is difficult to meet, here we develop an approach that allows us to implement (i) and address (ii). In particular, we used a computer graphics rendering approach (**Fig. 1**) to create stimuli for which we could independently manipulate monocular and binocular gloss cues.

In particular, we manipulated the rendering process to change the locations from which pixel intensities are determined while keeping the viewing position constant (see Muryy et al., 2014 for a detailed description). This allowed us to create four binocular conditions. First, we used physically correct rendering of objects with mirrored surfaces reflecting a natural scene (**Fig. 1B**: Mirror). Second, we created a 'painted' condition in which the reflections were 'stuck' onto the surface of the object. This had the effect that monocular features were almost identical to a glossy object, but when stimuli were viewed stereoscopically the object appeared matte (Muryy et al., 2013; see also Doerschner et al., 2011 for the analogous case with motion). Third, we modified the rendering process to create physically incorrect specular reflections (**Fig. 1B**: Anti-mirror). These stimuli had different overall disparity values, but nevertheless evoked an impression of surface gloss. Finally, we presented the same image to the two eyes, creating the impression of a stereoscopically flat object for which gloss was defined solely by monocular cues (**Fig. 1B**: Flat). We thereby sought to test for neural responses relating to changes in binocular signals vs. the perceptual interpretation of surface material properties. In addition, to draw comparisons with neuronal responses to gloss defined by monocular cues, we also measured fMRI responses when

participants viewed stimuli for which we used an image editing technique to alter the impression of surface gloss (**Fig. 1C**). In this way, we aimed to reveal common responses to gloss defined by differences in monocular and binocular cues.

Methods

Participants

Twelve participants with normal or corrected-to-normal vision took part in the experiment. One was an author H.-C. S. and the remainder were naïve. Three were male, and age ranged between 19 to 39 years. Participants were screened for normal stereoacuity and MRI safety. They provided written informed consent. All participants took part in three fMRI sessions: one binocular gloss session, one non-stereoscopic gloss session (see 'Stimuli' and 'Design and Procedure'), and one localiser session (see 'ROI definition'). The study was approved by the STEM Ethical Review Committee of the University of Birmingham. Non-author participants received course credits or monetary compensation.

Apparatus and Stimuli

Apparatus. Stimulus presentation was controlled using MATLAB (Mathworks Inc.) and Psychtoolbox (Brainard, 1997; Pelli, 1997). The stimuli were back projected by a pair of projectors (JVC DILA SX21) onto a translucent screen inside the bore of the magnet. To present stereoscopic stimuli, the projectors were fitted with spectral comb filters (INFITEC, GmbH) – see (Preston, Kourtzi, & Welchman, 2009). This presentation technique allows stereoscopic presentation of colour images, with only slight differences in the colour spectra-presented to each eye, and low cross-talk between the two eyes' views. Participants viewed the stimuli binocularly via a front-surface mirror fixed on the head coil with a viewing distance of 65 cm. In the non-stereoscopic gloss session, participants viewed stimuli (binocularly) without wearing the Infitec glasses. Luminance outputs from the projectors were measured using Admesy Brontes-LL colorimeter (Admesy, Netherlands) and then

147 linearized and equated for the RGB channels separately with Mcalibrator2 (Ban &
148 Yamamoto, 2013). Participant responses during the scan were collected using an optic fibre
149 button box.
150
151 *Stimuli.* A central fixation square (0.5 deg side length) was displayed in the
152 background to provide a constant reference to promote correct eye vergence. We performed
153 the experiment in two sessions: a binocular gloss session, and a non-stereoscopic gloss
154 session. For the binocular gloss session, we used Matlab to create three different 3D objects
155 ('potatoes', created by randomly distorted spheres which look like potatoes at arm's length
156 (Muryy et al., 2014; Muryy et al., 2013)). The rendering procedure involved using objects
157 with known surface geometries presented at a viewing distance of 65 cm (**Fig 1A**). The
158 objects had perfectly specular surfaces, and reflected one of three different spherical
159 illumination maps, that for rendering purposes were located at optical infinity (**Fig. 1A**). The
160 rendered images produced objects that were approximately 7 deg in diameter. These were
161 presented at the centre of the screen with ± 0.4 degree jitter from the centre to reduce the
162 build up of adaptation across repeated presentations at the centre of the screen.
163 To produce stimuli for the four experimental conditions ('mirror', 'painted', 'anti-
164 mirror', 'flat') in the binocular gloss session, we made subtle modifications to the stimulus
165 rendering process (for full details and mathematical implementation, see Muryy et al., 2014).
166 In particular, under standard mirror reflection (**Fig. 1B**: Mirror), stimuli are rendered by
167 finding the pixel value of point $\boldsymbol{P}$ in the image of left eye ($\boldsymbol{E_L}$) and right eye ($\boldsymbol{E_R}$) by reflecting
168 the viewing vectors from left eye ($\boldsymbol{V_L}$) and right eye ($\boldsymbol{V_R}$) around the surface normal, $\boldsymbol{n}$, to
169 calculate the reflected ray vectors $\boldsymbol{\omega_L}$ and $\boldsymbol{\omega_R}$ (e.g., $\boldsymbol{\omega_L} = 2\ (\boldsymbol{n}\ \boldsymbol{V_L}\ )\ \boldsymbol{n} + \boldsymbol{V_L}$). These point to
170 particular image intensities in the spherical illumination map, determining the pixel
171 intensities that should be presented to $\boldsymbol{E_L}$ and $\boldsymbol{E_R}$ (see cartoons next to stereo pairs in **Fig 1B**

for an illustration of this process). Using computer graphics, we changed the locations from which the objects are imaged for the purpose of defining the pixel intensities of the object, while keeping the stereoview frustum constant (**Fig. 1B**) (see Muryy et al. (2014)). This allowed us to manipulate the stereoscopic information from the reflections to create four different conditions, while leaving monocular images almost constant. Specifically:

First, in the 'mirror' condition (**Fig. 1B**: Mirror), stimuli are generated following the normal specular reflection, creating the impression of a mirrored object. Second, in the 'painted' condition (**Fig. 1B**: Painted), the specular reflections act like a texture and are effectively 'stuck' onto the surface of the object. This means that the specular reflections have the same stereoscopic depth as the object's surface, although the images still contain classic 'monocular' signals to reflection, such as the distortions of the surrounding illumination map. In the painted case, the stereoscopic information largely overrides these monocular cues, greatly reducing the perception of surface gloss (**Fig 2**). Third, in the 'anti-mirror' condition (**Fig. 1B**: Anti-mirror), we reversed the locations from which image intensities in the environment are determined between the two eyes. This leads to a considerable change the disparity structure of the images (Muryy et al., 2013); nevertheless the stimuli are perceived to have a similar glossy appearance to that of a correctly-rendered mirror (Muryy et al., 2012) (**Fig 2**). Finally, we created a 'flat' condition (**Fig. 1B**: Flat) in which the same image of the object was presented to both eyes, again reducing participants' overall impression of gloss (**Fig. 2**).

To ensure generality in identifying signals related to surface appearance, we used a different set of stimuli in the non-stereoscopic gloss session. In particular, we used single view renderings of 3D objects (3 different shapes) generated in Blender 2.67a (The Blender project: http://www.blender.org/). Participants were presented stimuli in four conditions (Glossy, Matte, Rough and Textured, see (Sun, Welchman, Chang, & Di Luca, 2016)). Only

data from the Glossy and Matte conditions are presented here. The Rough and Textured conditions are not directly relevant to the current study. To generate the Glossy and Matte stimuli, we first rendered the objects with a specular surface component. We then edited the images in Adobe Photoshop, using the 'color range' tool to extract the portions of the objects corresponding to specular reflections (i.e., lighter portions of the shape in **Fig. 1C**, where *fuzziness* parameter of the color range tool was set to 40 to isolate the specular highlights). We then pasted these highlights onto a rendering of the object produced with no specular surface reflection. When pasted into the 'correct' locations (i.e. those that contained highlights for the specular surface) the object appeared glossy (**Fig 1C**, top); however, when rotated 45 deg in the image plane, the surface no longer appeared glossy (**Fig 1C**, bottom) (Wilcoxon signed-rank test, two-tailed, N=7, Z=-2.2, p<.05). This difference in appearance between the two conditions is likely to be due to the incoherence between the position/orientation of the highlights and the contextual information about shape and illumination (Anderson & Kim, 2009; Kim et al., 2011; Marlow et al., 2011).

Note that the basic appearance of the stimuli is (deliberately) quite different for the binocular (**Fig. 1B**) and non-stereoscopic (**Fig. 1C**) imaging sessions, as we wished to test for generalisation of the impression of gloss that could not be ascribed to simple image features (e.g. contours) or the overall 3D shape. Moreover, note that we did not directly compare brain activity between the two types of stimuli – rather we looked for generalisation across contrasts conducted within each stimulus set (i.e., 'gloss vs. matte' generalised to 'mirrored vs. painted').

MRI data acquisition. A 3-Tesla Philips Achieva scanner with an 8-channel phase-array head coil was used to obtain all MRI images at the Birmingham University Imaging Centre (BUIC). Functional whole brain scans with echo-planar imaging (EPI) sequence (axial 32 slices, TR 2000 ms, TE 35 ms, voxel size 2.5 × 2.5 (inplane) × 3 (thickness) mm, flip

angle 80 deg, matrix size 96 × 94) were obtained for each participant. The EPI images were acquired in an ascending interleaved order for all participants. The same sequence was used in both sessions. T1-weighted high-resolution anatomical scans (sagittal 175 slices, TR 8.4 ms, TE 3.8 ms, flip angle 8 deg, voxel size: 1 mm^3) were also obtained to reconstruct cortical surfaces of individual participants and to achieve precise co-registrations of EPI images onto individual anatomical spaces.

Design and Procedure

A block design was used in both sessions. Each session took about 1.5 hours during which each participant completed in 7 to 10 runs for binocular gloss session and 8 to 10 runs for non-stereoscopic gloss session (depending on setup time and the participants' needs to rest between scans). The run length was 400 s and 368 s for the binocular- and non-stereoscopic- gloss sessions, respectively. Each run started with four dummy scans to prevent startup magnetization transients and consisted of 16 experimental blocks each lasting 16 s. There were 4 block types (i.e., one for each condition), repeated four times in a run. In each block of the binocular gloss session 10 objects were presented in a pseudo-random order. Stimuli were presented for 1000 ms with 600 ms interstimulus interval (ISI). Participants were instructed to maintain fixation and perform an oddball task for glossiness judgments. Specifically, at the end of each block (signalled to the participants by a change in the fixation marker) participants had to indicate if all of the presented objects had the same glossiness (i.e., all matte, or all glossy), or whether one of the presented objects differed in gloss. They had two seconds to make their response before the next block began. They were able to perform this task well (mean d'=2.04; SEM=0.31). Five 16 s fixation blocks were interposed after the third, fifth, eighth, eleventh, and thirteenth stimulus blocks to measure fMRI signal baseline. In addition, 16 s fixation blocks were interposed at the beginning and at the end of

the scan, making a total of seven fixation blocks during one experimental run. An illustration of the scan procedure is provided in **Fig. 3**. In the non-stereoscopic gloss session, stimuli were presented for 500 ms with 500 ms ISI. Participants were instructed to maintain fixation and perform a 1-back matching task, whereby they pressed a button if the same image was presented twice in a row. They were able to perform this task well (mean d'=2.03; SEM=0.10). Other details were the same as for the binocular gloss session.

Data analysis

Functional MRI data processing. The basic data processing procedures for both the binocular and the non-stereoscopic gloss sessions are identical to our previous studies (Sun et al., 2015; Sun et al., 2016). To summarise the procedure, we computed the global signal variance of the blood oxygenation level dependent (BOLD) signal for each run using the whole-brain average of activity across volumes. If this exceeded 0.23% the scan run was excluded from further analysis to avoid the influence of scanner drifts, physiological noise or other artifacts (Junghöfer, Schupp, Stark, & Vaitl, 2005). On this basis, 17/146 runs and 6/118 runs across 12 participants for binocular and non-stereoscopic gloss session respectively were excluded from further analysis.

ROI definition. A total of 15 regions of interest (ROIs) were defined. For all participants V1, V2, V3v, V4 V3d, V3A, V3B/KO (kinetic occipital region), hMT+/V5 (human motion complex), LO (lateral occipital region) and pFs (posterior fusiform sulcus) were defined by localizers in a separate session as in previous studies (Ban, Preston, Meeson, & Welchman, 2012; Dövencioğlu, Ban, Schofield, & Welchman, 2013; Murphy, Ban, & Welchman, 2013; Sun et al., 2015). For 7 of the 12 participants, higher dorsal areas V7, ventral intraparietal sulcus (VIPS), parieto-occipital IPS (POIPS), dorsal IPS medial (DIPSM), and dorsal IPS anterior (DIPSA) were also defined by a localizer in which random-

dot stereogram with 3D structure from motion (SfM) information was contrasted with moving dots without stereogram and SfM information (Orban et al., 2006; Orban, Sunaert, Todd, Van Hecke, & Marchal, 1999). For the other 5 participants, V7 was identified as anterior and dorsal to V3A and other dorsal areas defined according to Talairach coordinates ([x,y,z] = [30, -78, 27] for right VIPS; [-27, -72, 30] for left VIPS; [24, -75, 45] for right POIPS; [-18, -72, 54] for left POIPS; [18, -60, 63] for right DIPSM; [-15, -63, 60] for left DIPSM; [39, -36, 54] for right DIPSA; [-36, -48, 60] for left DIPSA) and draw around GLM t-value maps that had *t* value greater than zero for the contrast of "all experiment conditions vs. fixation block" (Dövencioğlu et al., 2013; Murphy et al., 2013; Orban et al., 2003).

Additional fMRI analysis. We used multivoxel pattern analysis (MVPA) to compute prediction accuracies for the experimental conditions. We selected voxels by first computing the contrast "all experimental conditions vs. fixation", and then selecting the top 250 voxels from this contrast within each ROI of each individual participant (Ban et al., 2012). If a participant had fewer than 250 voxels in a particular ROI, we used the maximum number of voxels that had t values greater than 0. After selecting the voxels, we extracted the time series (shifted by 4 s to account for the hemodynamics response delay) and converted the data z-scores. Then, the voxel-by-voxel signal magnitudes for a stimulus condition were obtained by averaging over 8 time points (TRs) (= 1 block) separately for each scanning run. To remove baseline differences in the response patterns between stimulus conditions and scanning runs, we normalized by subtracting the mean for each time point. To perform the multi-voxel pattern analysis (MVPA), we used a linear Support Vector Machine (SVM) implemented in libsvm toolbox (http://www.csie.ntu.edu.tw/~cjlin/libsvm, Chang & Lin, 2011) to discriminate the different conditions in each ROI. In the training phase, 24 response patterns for each stimulus condition were used as a training dataset for those participants that completed 7 runs and 36 response patterns were used for those who completed 10 runs. Then,

4 response patterns for each condition were classified by the trained classifier in the test phase. These training/test sessions were repeated and validated by a leave-one-run-out cross-validation procedure. The ROI-based prediction accuracy for each participant was defined as a mean of these cross-validation classifications. In situations where there were different numbers of samples between two conditions in a contrast (e.g., mirror & anti-mirror vs. painted), we used balanced weight vectors for each class by adjusting the *j* parameter in libsvm toolbox to eliminate bias from different number of samples in the training dataset. We also used a searchlight classification analysis approach (Kriegeskorte, Goebel, & Bandettini, 2006) whereby we defined a spherical ROI with 8 mm radius, and moved it through the entire volume of cortex with masking volumes so that the searchlight sphere only captured gray matter voxels. For each location, we recomputed the SVM classification analysis.

Results

To test for visual responses related to binocular and monocular cues to gloss, we first identified regions of interest within the visual and parietal cortex using independent localizer scans (**Fig. 4**). We then used MVPA to test for responses related to the impression of glossy *vs*. matte surfaces. In particular, we used responses in different experimental conditions to understand how fMRI signals might relate to changes in the material appearance of the viewed object vs. changes in the disparity-defined depth structure. To this end, we concentrated on three main contrasts (**Fig. 5A**). First, we tested for responses related to surface gloss, contrasting the mirror and anti-mirror conditions (both perceived as glossy, **Fig 2**, and their averaged overall disparity is (approximately) the same as in painted condition) against the painted object (perceptually matte). Second, we performed a contrast between the mirror and anti-mirror conditions; the logic of this contrast is that while both appear glossy, the raw disparity composition of the shapes is quite different. Third, we contrasted the painted and flat conditions, which provides the maximal change in 3D shape, while both are interpreted as not evoking a strong impression of gloss (**Fig. 2**). In the extreme scenario of a cortical region specialized for processing surface material, we would expect to be able to decode glossy vs. matte renderings of the stimuli, but not the difference between mirror and anti-mirror conditions, or the difference between the painted and flat conditions.

We found that we were able to predict the stimulus from the fMRI data at levels reliably above chance ($p<.05$, one-tailed, Bonferroni corrected) in multiple regions of interest (V4, LO, V3d, V3A, V3B/KO, hMT+/V5, V7, VIPS, DIPSM, DIPSA) when contrasting the mirror and anti-mirror conditions against their painted counterparts (**Fig. 5A**, black data series). This suggests widespread sensitivity to differences in the material that comprise the stimuli whether the specular reflections are physically correct or not. Considering the

differences between the mirror and anti-mirror conditions (**Fig. 5A**, gray bars), we were not able to reliably predict the stimuli in any regions of interest. This failure to decode differences between the two conditions might suggest widespread responses that respond to glossy appearance, and thus do not differentiate between the mirror and anti-mirror conditions. Nevertheless, interpreting such a null result requires caution: disparity differences between the stimuli may have been insufficient to support decoding, or the size of the differences between mirror and anti-mirror conditions may have been dwarfed by the disparity differences between the different 3D shapes that were presented. Finally, contrasting the painted and flat conditions (**Fig. 5A**, white bars) revealed above chance prediction accuracies in V3B/KO, hMT+/V5, V7 and LO ($p<.05$, one-tailed, Bonferroni corrected). Decoding performance in this condition allows us to identify areas sensitive to changes in the 3D structure of the shapes. The result is consistent with previous work suggesting sensitivity to disparity-defined depth in these areas (Ban et al., 2012; Dövencioğlu et al., 2013; Murphy et al., 2013).

To facilitate comparison of performance between conditions, we calculated a '3D structure index' to examine decoding performance that could be attributed to information about 3D shape. We expressed prediction performance in units of discriminability (d') and contrasted performance for the mirror vs. anti-mirror condition with the painted vs. flat condition based on a simple subtraction. The logic of this contrast is that for both sets of comparisons there is minimal difference in the material appearance of the shapes, so the contrast reflects differences in the 3D structure of the shapes in both conditions. We also created a 'Gloss index' by contrasting performance in the mirror vs. anti-mirror contrast with the [mirror and anti-mirror] vs. painted classification. The logic of this contrast is to compare similarly glossy objects (with different disparity information) against differentially glossy

objects (with different disparity information). The formulas of the two indices are presented as below:

3D structure index = d'(painted vs. flat) – d'(mirror vs. anti-mirror)

Gloss index = d'(mirror & anti-mirror vs. painted) – d'(mirror vs. anti-mirror)

We used mirror vs. anti-mirror as baseline for normalizing 3D structure index and Gloss index because in this contrast both conditions have the same visual appearance (glossy) and similar 3D structure. The comparison between the two indices is suggestive of whether a brain area is more specialized for gloss processing or 3D structure processing. We present the two indices across all ROIs in **Fig. 5B**. We first considered whether the indices are significantly above chance level (p<.05, one-tailed, Bonferroni corrected), using permutation tests to calculate 95% shuffled baseline of d' difference for Gloss index (0.14) and 3D structure index (0.16). We found that the Gloss index was significantly above chance in DIPSA (t_{11}=4.4, p<.01) and LO (t_{11}=5.3, p<.01), suggesting that signals in these areas are discriminable based on gloss information. For the 3D structure index, we found sensitivity significantly above chance in V3B/KO (t_{11}=3.5, p<.05) and LO (t_{11}=4.1, p<.05). These results suggest that LO processes information relevant to both 3D structure and material properties.

We next sought to compare the indices against each other. To this end we ran a 2 (Gloss index and 3D structure index) × 15 (ROIs) repeated-measures ANOVA. This indicated a main effect of ROI ($F_{14,154}$=2.5, p<.01) and importantly a significant interaction with index ($F_{14,154}$=2.8, p<.01). We then used post-hoc contrasts to test the differences between the indices in each ROI. We found a significantly higher Gloss index in V2, pFs, DIPSM and DIPSA, suggesting areas preferentially engaged in the processing of material properties (**Fig. 5B**, asterisks indicate p<.05). It is reassuring to note that areas V2 and pFs

were previously found to be involved in the processing of information about specular reflectance from monocular cues (Sun et al., 2015; Wada et al., 2014), suggesting that they represent general information about surface gloss regardless of the source. In summary, LO appears to process both surface properties and 3D structure information, while V2, pFs, DIPSM and DIPSA selectively process surface properties. Transfer analysis between [mirror & anti-mirror *vs.* painted] and [flat *vs.* painted] suggested that the processing of surface properties and 3D structure information involves in the same voxels in LO (see **Fig. 6**).

To ensure that we had not missed any important loci of activity related to gloss or structure, we used a searchlight classification analysis (**Fig. 4A**). This confirmed that locations identified by the searchlight procedure fell within those we had sampled using our region of interest localiser approach.

In addition to making measurements of binocularly-defined gloss, we also used an image editing procedure to alter the impression of gloss evoked by monocular cues (**Fig. 1C**). As an initial analysis of the fMRI responses evoked by viewing these stimuli, we tested for the ability of an MVPA classifier to discriminate glossy vs. matte stimuli. **Fig. 7A** shows the classification results of Glossy *vs.* Matte stimuli. We found widespread performance above chance ($p<.05$, one-tailed, Bonferroni corrected) when comparing between glossy and matte versions of the stimuli (V1, V2, V3v, V4, LO, pFs, V3d, V3A, V3B/KO, POIPS). This was consistent with an MVPA of data collected in a previous study (Sun et al., 2015) that contrasted objects rendered with different surface reflection parameters to alter perceived gloss (**Fig. 7B**). This also indicates that the additional conditional conditions (Rough and Textured) that were used in the non-stereoscopic gloss session had a very limited effect on gloss processing, because the results are consistent with our previous study (Sun et al., 2015) did not contain Rough and Textured conditions. We also used a searchlight procedure for

these data (**Fig 4B**), that confirmed we had captured relevant responses within our independently-localised ROIs.

Considering the non-stereoscopic gloss results together with the preceding binocular gloss results, suggests that some cortical areas (i.e. V3d, V3A, V3B/KO, V4, LO) support the decoding of both monocular and binocular gloss cues. However, our critical interest was whether the same neural populations (as sampled by voxels) were involved in processing of both binocular and monocular gloss cues. To examine this issue, we performed a transfer analysis to test whether training a classifier on gloss defined by monocular cues (non-stereoscopic imaging session) would support predictions for fMRI responses evoked by binocular cues (and vice versa). Our expectation was that a cortical area that shows transfer in both directions would suggest an area intricately involved in processing gloss, regardless of its image source.

We first trained the SVM classifier to discriminate Glossy *vs.* Matte conditions in the non-stereoscopic gloss session and then tested whether the classifier could discriminate [mirror & anti-mirror] *vs.* painted activation in the binocular gloss session. We found significant transfer from monocular to binocular gloss in areas V1, V2, V3d and V3B/KO (**Fig. 8**: black bars). We then tested whether there was transfer from binocular gloss to monocular gloss, but found no evidence for transfer in this direction (**Fig. 8**: white bars). As a follow-up analysis, we also conducted a searchlight classification analysis, in case our ROI approach did not capture important loci of activity. This analysis confirmed our choice of ROIs, and reconfirmed that while we observed transfer from monocular to binocular gloss cues (**Fig. 9A**), we did not observe transfer from binocular to monocular gloss cues (**Fig. 9B**).

General discussion

Here we sought to test for cortical areas involved in the processing of gloss from binocular and monocular cues to surface material. We sampled fMRI activity from across the visual processing hierarchy, and contrasted fMRI responses in conditions that evoked different impressions of surface gloss. We found that ventral area LO supported the decoding of information about both (i) the material properties of objects and (ii) 3D structure. By contrast we found that differences in gloss were more discriminable than differences in disparity-defined shape based on fMRI responses in dorsal anterior Interparietal Sulcus (DIPSA). We contrasted responses to monocular and binocular signals to gloss, finding differential involvement of areas within the dorsal and ventral streams. Importantly, V3B/KO appeared to be involved in the processing of both types of information. This was supported by a transfer analysis that showed binocularly specified gloss could be decoded using an algorithm trained on differences in perceived gloss specified by monocular features. These results point to the involvement of both ventral and dorsal brain areas in processing information related to gloss, with an intriguing confluence in area V3B/KO that has previously been associated with the processing of 3D structure.

Our approach to investigating binocular cues to gloss was to make subtle modifications to the rendering process so that low-level image statistics were almost identical between different conditions. This allowed us to test for the neural processing of binocular signals to surface reflectance properties, which are likely to interact with the processing of monocular cues to gloss (such as the luminance intensity of specular reflections, their contrast and spatial frequency (Marlow & Anderson, 2013; Marlow, Kim, & Anderson, 2012; Motoyoshi et al., 2007; Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008). To test the impression of gloss from monocular cues, we also used a simple image-editing technique that altered participants' impressions of gloss by rotating specular highlight components in the

image plane. This broke the relationship between surface curvatures specified by the image and the location of reflections (**Fig. 1C**) and ensured that low-level image features were near-identical (Anderson & Kim, 2009; Kim et al., 2011; Marlow et al., 2011). This is a different procedure to that used in previous studies that used spatial scrambling, phase scrambling or changing overall luminance (Okazawa et al., 2012; Sun et al., 2015; Wada et al., 2014). It is reassuring that the results of this manipulation (**Fig. 7A**) converge with a comparable analysis of results from a previous study that employed image scrambling (**Fig. 7B**) (Sun et al., 2015) . In particular, both data sets indicate that monocular gloss cues are processed in ventral areas as well as in dorsal areas V3d, V3A and V3B/KO.

More broadly, our results suggest that gloss-related signals are processed in earlier visual areas (V1, V2, V3d, V3v) and ventral visual areas (V4, LO, pFs), consistent with previous findings (Okazawa et al., 2012; Wada et al., 2014). We provide converging evidence in line with two previous studies (using a different approach to generate stimuli) that human V3B/KO is involved in gloss processing (Sun et al., 2015; Wada et al., 2014). In addition, our results indicate that higher dorsal area POIPS supports the decoding of monocular gloss cues (**Fig. 7A**). This is not something that has been found before (Sun et al., 2015; Wada et al., 2014). It is possible that our use of MVPA to analyse these data provides a more sensitive tool to reveal representations that were not detected using the standard general linear model contrasts in previous work. However, it is also possible that our image editing technique evoked the impression of surface occlusion that increased the complexity of the viewed shape, and may have promoted subtle differences in the degree to which the stimuli engaged the participants' attention.

It is informative to compare the results we obtained in the non-stereoscopic and binocular gloss imaging sessions. Results from the non-stereoscopic gloss manipulations indicated responses in V1 and V2 that were not identified in by the binocular gloss

manipulations: this may be due to the very strong image similarity of the images across conditions for the binocular stimuli (**Fig. 1B**). In contrast, dorsal area V3d, V3A and V3B/KO were found respond to both monocular and binocular gloss cues. This pattern suggests that these areas may represent general information about surface gloss regardless of how it is conveyed. Other dorsal areas (especially for hMT+/V5, V7, VIPS DIPSM and DIPSA) were engaged by the binocular gloss information but not by monocular gloss cues. Our finding of this dorsal involvement was not anticipated from previous studies of material perception; however, it is broadly consistent with previous imaging studies that have pointed to the strong involvement of dorsal areas in processing binocular cues (Ban et al., 2012; Dövencioğlu et al., 2013; Murphy et al., 2013; Neri, Bridge, & Heeger, 2004; Vanduffel et al., 2002). Higher ventral areas such as V4 and LO were also found to be involved in processing binocular gloss information. This is compatible with previous fMRI studies of material perception that have pointed to the involvement of higher ventral areas (Cant & Goodale, 2007, 2011; Cavina-Pratesi, Kentridge, Heywood, & Milner, 2010a, 2010b; Hiramatsu, Goda, & Komatsu, 2011).

It is important to note that slightly different experimental procedures and tasks were used for the binocular and non-stereoscopic gloss sessions. In particular, we used an oddball task for binocular session to make participants focus on binocular gloss information instead of simply judging on monocular changes (i.e. illumination and object shape), while we used a 1-back task in non-stereoscopic session. These differences may have affected the difference of SVM classification performance between the two sessions. However, the performance difference across ROIs within each session should not have been affected. Moreover, the evidence of transfer in V3B/KO despite differences in procedure may offer reassurance that this result is likely to be due to the common factors (i.e., gloss) between experiments, rather than differences in task or the 3D shapes.

While we found clear evidence for fMRI responses that differentiated glossy and non-glossy binocular cues, we did not find activity patterns that supported the decoding of mirror vs. anti-mirror stimuli. From the perspective of the impression of surface material this is not surprising (these stimuli look equally glossy), however, the stimuli do contain differences in binocular disparities that we might expect the brain to be able to decode. Nevertheless, our stimuli contained disparities that are difficult to fuse (Muryy et al., 2014), perhaps leading to unstable and/or unreliable estimates of binocular disparities. In addition, we presented different shapes that had different disparity structures, meaning that the disparity differences within a shape between mirror and anti-mirrored stimuli may have been overcome by the differences between individual shapes.

We found that the preference for processing information about binocular gloss vs. 3D structure differed across ROIs. In particular, we found that V3B/KO, hMT+/V5, V7 and LO not only responded to binocular gloss information but also information about 3D structure (**Fig. 5A**). The comparison between the Gloss index and the 3D structure index (**Fig. 5B**) shows that V2, DIPSM, DIPSA and pFs had better classification performance for decoding binocular gloss information than 3D structure information, indicating that these areas may be more specialized for processing surface properties than 3D structure. Interestingly, V2 and pFs were also found to have selectivity for gloss information from specular reflectance in previous studies (Okazawa et al., 2012; Wada et al., 2014) as well as in the current study (**Fig. 7**). The relatively weaker decoding performance in V2 and pFs for binocularly defined gloss suggests a preference for monocular gloss cues in these areas. By contrast, LO appears to respond to information about binocular gloss and 3D structure equally well (**Fig. 5B**) and most importantly, it was the only ROI that showed strong transfer effect between the two kinds of information (**Fig. 6**). One possible explanation is that the processing of binocular

gloss and 3D structure influence each other, as shown by previous psychophysical studies (Blake & Bülthoff, 1990; Muryy et al., 2013).

A direct way to examine whether an area combines monocular and binocular gloss cues and represents surface gloss in a general way is to test whether the activities that afford classification evoked by one cue type can transfer to the classification of the other. Here we trained an SVM classifier to discriminate between glossy and matte objects for monocular and binocular gloss information and found transfer effects from monocular to binocular cues in left V3B/KO (as well as a small part of V3v and V1, see **Fig. 9**). However, we did not find a transfer effect from binocular to monocular gloss cues. A possible explanation for this asymmetry is that the underling neural populations that respond to binocular gloss are more specialised than those that respond to monocular gloss. Under this scenario, we would conceive that a relatively large population of neurons responds to monocular gloss cues, but only a subset of these neurons responds to both monocular and binocular cues. When the classifier is trained on binocular differences, it would select the units that respond to both cues. However, a classifier trained on monocular gloss differences could select voxels reflecting a broad population, many of which do not respond to binocular cues.

More generally, this architecture might suggest that the neural representation of surface material involves a number of co-localised but specialist neuronal populations that respond to a range of different cues that are diagnostic of surface gloss. Previous studies have identified various monocular cues that could contribute to the perception of gloss (Anderson & Kim, 2009; Doerschner et al., 2011; Fleming et al., 2003; Gegenfurtner et al., 2013; Kim & Anderson, 2010; Kim et al., 2011; Kim et al., 2012; Landy, 2007; Marlow & Anderson, 2013; Marlow et al., 2011; Motoyoshi et al., 2007; Nishio et al., 2012; Nishio et al., 2014; Okazawa et al., 2012; Olkkonen & Brainard, 2010; Sun et al., 2015; Wada et al., 2014) and discussed in detail the computations involved in decomposing the intensity gradients in images of

surfaces into distinct causes (shading, texture markings, highlights, etc.). Each of these subtypes may be encoded by specialist populations whose aggregated effect supports the impression of gloss. In the case of the binocular gloss cues we have studied, it seems likely that the brain exploits information about image locations that are difficult to fuse due to large vertical (ortho-epipolar) disparities or horizontal (epipolar) disparity gradients whose magnitude exceed fusion limits (Muryy et al., 2013). One means of conceptualising the differences between the binocular stimuli we used is in terms of the complexity of the binocular disparity signals – i.e., mirror and anti-mirror stimuli could be thought of as more complex (because of the large disparities) than the painted and flat stimuli. Our results suggest differences between these conditions that align to differences in the perceptual impression of gloss; however, we cannot rule out that the critical differences related to overall disparity complexity *per se* rather than gloss. Under this scenario, the areas we have localised might correspond to a halfway house between a metric based on complexity and one based on the appearance of gloss. Nevertheless, our observation of transfer between monocular and binocular gloss cues is suggestive of a representation of gloss *per se*.

In summary, here we used systematic manipulation of binocular gloss cues to test for cortical areas that respond to surface material properties. We show the involvement of regions within the ventral and dorsal streams, and draw direct comparisons with cortical responses defined by monocular gloss cues. Our results point to the potential integration of binocular and monocular cues to material appearance in area V3B/KO that showed partial evidence for transfer between different signals.

Acknowledgements

This project was supported by fellowships to A.E.W. from the Wellcome Trust (095183/Z/10/Z) and to H.B. from the Japan Society for the Promotion of Science (JSPS KAKENHI (26870911)).

Figure captions

Figure 1. Stimuli used for binocular and non-stereoscopic gloss experiments. (A) Synthetic objects ("potatoes") were rendered under three different illumination maps to create the stimuli. (B) Schematic illustration of the rendering procedure and example stereograms for each condition (cross the eyes to fuse the image pairs). Mirror condition: reflections entering each eye follow the law of specular reflection, creating a physically correct image of a polished object reflecting its surrounding environment (schematically illustrated using the color spectrum). Painted condition: pixel intensities for each location on the surface of the object are determined based on the reflection of a ray cast from midway between the participant's eyes. The object is imaged from the true positions of the two eyes, meaning that the environment effectively acts as a texture painted onto the surface of the object. Anti-mirror condition: the reflected ray vectors are reversed for the two eyes, so the left eye images a portion of the environment appropriate for the right eye. This alters the disparities produced by reflection, but the object appears glossy. Flat condition: we randomly select the image of one eye (the right eye in the example) and present it to both eyes. Objects look flat and made specular reflections have the same apparent depth as the image plane. (C) An example stimulus in non-stereoscopic gloss session. Specular components are presented in Glossy condition while in the Matte condition the specular components are rotated by 45 degrees in the image plane, making the object appear matte.

Figure 2. Results of psychophysical ratings of perceived gloss for the different binocular conditions. Participants (n=6, different from the participants of scan sessions) were presented with four pairs of stereo stimuli (corresponding to the four conditions) concurrently on a

610 screen viewed with 3D prism glasses (NVP3D) in the laboratory. The shape and illumination
611 of each stimulus pairs were randomly chosen from the three different potato shapes and the
612 three different illumination maps described in **Fig. 1**. Participants were asked to choose (i) the
613 most and (ii) the least glossy object by pressing numerical keys which correspond to the
614 position of the four stereo stimuli on the screen. Judgements were blocked into 180 trials,
615 with block order counterbalanced across participants. The probability of choosing each
616 condition was averaged across participants. Bar graphs show mean selection probability ± 1
617 SEM. A one-way repeated-measures ANOVA (mirror, painted, anti-mirror, flat) was
618 significant for both blocks ($F_{3,15}$=12.0, p<.001 for most glossy block; $F_{3,15}$=27.3, p<.001 for
619 least glossy block). Asterisks represent significant differences based on Tukey's HSD post
620 hoc tests (p<.05).
621
622 **Figure 3.** The stimulus presentation protocol in binocular gloss session for one scan. On each
623 run, 23 blocks were presented (16 s + 2 s response time each), including 7 fixation blocks and
624 16 experimental blocks. During each experimental block, stimuli were presented for 1000 ms
625 with 600 ms interstimulus interval (ISI). Participants were instructed to detect stimuli that
626 differed from the others in terms of glossiness (oddball detection task for glossiness).
627
628 **Figure 4.** Searchlight classification analysis results for binocular (A) and non-stereoscopic
629 (B) gloss conditions across 12 participants. The color code represents significant t-value of
630 mirror *vs.* flat and Glossy *vs.* Matte classification accuracies in (A) and (B) respectively
631 (testing against chance level 0.5). Blue dashed lines are the ROI boundaries we defined with
632 independent localizer scans. The significance level is p<.05 with cluster-size thresholding
633 25mm^2. Regions with significant results are presented on the flat maps of one representative
634 participant. Note that since classification results are averaged across participants and then

presented on the flat maps of one representative participant, individual ROI boundaries may not perfectly fit the group level.

Figure 5. MVPA prediction performance across 12 participants for (A) [mirror & anti-mirror] *vs.* painted (black bars), mirror *vs.* anti-mirror (gray bars), and painted *vs.* flat (white bars). The bars reflect mean prediction accuracy with ± 1 SEM. Solid horizontal lines represent chance performance for the binary classification (0.5). Dotted horizontal lines represent the upper 95th percentile with Permutation tests (1000 repetitions for each ROI of each participant with randomly shuffling stimulus condition labels per test. The one-tailed, 95% boundaries of accuracy distributions were averaged across all ROIs, which was 52.52% for [mirror & anti-mirror] *vs.* painted, 53.11% for mirror *vs.* anti-mirror, and 53.13% for *vs.* painted flat). Asterisks in the bottom of the bars represent accuracies significantly above the shuffled baseline ($p<.05$, one-tailed, Bonferroni corrected). (B) d' difference between [mirror & anti-mirror *vs.* painted] and [mirror *vs.* anti-mirror] classification is used as a Gloss index. The d' difference between [painted *vs.* flat] and [mirror *vs.* anti-mirror] is used as a 3D structure index. Dotted horizontal lines represent the upper 95th percentile of a permutation tests (1000 repetitions). Asterisks at the bottom of the bars indicate that the index was significantly above the shuffled baseline ($p<.05$, one-tailed, Bonferroni corrected). Black dots above bar pairs represent significant difference between the two indexes (Tukey's HSD post-hoc test at $p<.05$).

Figure 6. MVPA prediction performance across 12 participants for transfer analysis between [mirror & anti-mirror *vs.* painted] and [flat *vs.* painted]. We trained the SVM classifier to discriminate mirror & anti-mirror *vs.* painted and tested whether it is distinguishable for flat *vs.* painted (black bars). We also tested the transfer effect in the other way (white bars). The

bars reflect mean classification accuracy with ± 1 SEM. Solid horizontal lines represent chance performance 0.5 for the binary classification. Dotted horizontal lines represent the upper 95th percentile with Permutation tests (1000 repetitions. The one-tailed, 95% boundaries of accuracy distributions for black bars was 52.24% and 53.17% for white bars). Asterisks in the top of the bars represent that the accuracies were significantly above the shuffled baseline ($p<.05$, one-tailed, Bonferroni corrected).

Figure 7. MVPA prediction performance for Glossy *vs.* Matte in non-stereoscopic gloss session in the current study (A) and in our previous study (Sun et al., 2015) with a group of 15 participants (B). The bars reflect mean classification accuracy with ± 1 SEM. Solid horizontal lines represent chance performance 0.5 for the binary classification. Dotted horizontal lines represent the upper 95th percentile with Permutation tests (1000 repetitions. The one-tailed, 95% boundaries of accuracy distributions in A was 52.79% and 52.39% in B). Asterisks in the top of the bars represent that the accuracies were significantly above the shuffled baseline ($p<.05$, one-tailed, Bonferroni corrected). Higher dorsal areas (V7-DIPSA) were not defined in (B) as parietal localizer was not applied in that study.

Figure 8. (A) MVPA prediction performance across 12 participants for the transfer analysis between binocular and monocular gloss cues. We trained the SVM classifier to discriminate Glossy *vs.* Matte conditions in non-stereoscopic gloss session and tested whether it could predict [mirror & anti-mirror] *vs.* painted in the binocular gloss session (black bars). We also tested the transfer effect in the other way (white bars). The bars reflect mean classification accuracy with ± 1 SEM. Solid horizontal lines represent chance performance 0.5 for the binary classification. Dotted horizontal lines represent the upper 95th percentile with

permutation tests (1000 repetitions for each ROI). Asterisks above the bars represent that the accuracies were significantly above shuffled baseline ($p<.05$, one-tailed, without correction).

Figure 9. Searchlight transfer analysis results. In (A) we trained the SVM classifier to discriminate Glossy vs. Matte conditions in the non-stereoscopic gloss session and then tested [mirror & anti-mirror] vs. painted in the binocular gloss session. In (B) we tested for transfer in the opposite direction. The color code represents the t-value against chance level (0.5) with 25mm^2 cluster-size thresholding. Significant transfer is found primarily by training on non-stereoscopic gloss cues and subsequently testing on binocular information, but not in the opposite direction.

References

Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology, 21*(24), R978-R983. doi:10.1016/j.cub.2011.11.022

Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision, 9*(11). doi:10.1167/9.11.10

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature Neuroscience, 15*(4), 636-643. doi:10.1038/nn.3046

Ban, H., & Yamamoto, H. (2013). A non–device-specific approach to display characterization based on linear, nonlinear, and hybrid search algorithms. *Journal of Vision, 13*(6), 20-20. doi:doi: 10.1167/13.6.20

Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature, 343*(6254), 165-168. doi:10.1038/343165a0

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436. doi:10.1163/156856897X00357

Cant, J. S., & Goodale, M. A. (2007). Attention to Form or Surface Properties Modulates Different Regions of Human Occipitotemporal Cortex. *Cerebral Cortex, 17*(3), 713-731. doi:10.1093/cercor/bhk022

Cant, J. S., & Goodale, M. A. (2011). Scratching Beneath the Surface: New Insights into the Functional Properties of the Lateral Occipital Area and Parahippocampal Place Area. *The Journal of Neuroscience, 31*(22), 8248-8258. doi:10.1523/jneurosci.6113-10.2011

Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2010a). Separate Channels for Processing Form, Texture, and Color: Evidence from fMRI Adaptation

and Visual Object Agnosia. *Cerebral Cortex, 20*(10), 2319-2332. doi:10.1093/cercor/bhp298

Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2010b). Separate Processing of Texture and Form in the Ventral Stream: Evidence from fMRI and Visual Agnosia. *Cerebral Cortex, 20*(2), 433-446. doi:10.1093/cercor/bhp111

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 1-27. doi:10.1145/1961189.1961199

Doerschner, K., Fleming, Roland W., Yilmaz, O., Schrater, Paul R., Hartung, B., & Kersten, D. (2011). Visual Motion and the Perception of Surface Material. *Current Biology, 21*(23), 2010-2016. doi:10.1016/j.cub.2011.10.036

Doerschner, K., Maloney, L. T., & Boyaci, H. (2010). Perceived glossiness in high dynamic range scenes. *Journal of Vision, 10*(9). doi:10.1167/10.9.11

Dövencioğlu, D., Ban, H., Schofield, A. J., & Welchman, A. E. (2013). Perceptual integration for qualitatively different 3-d cues in the human brain. *Journal of Cognitive Neuroscience, 25*(9), 1527-1541. doi:10.1162/jocn_a_00417

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision, 3*(5). doi:10.1167/3.5.3

Gegenfurtner, K., Baumgartner, E., & Wiebel, C. (2013). The perception of gloss in natural images. *Journal of Vision, 13*(9), 200. doi:10.1167/13.9.200

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *NeuroImage, 57*(2), 482-494. doi:10.1016/j.neuroimage.2011.04.056

751 Junghöfer, M., Schupp, H. T., Stark, R., & Vaitl, D. (2005). Neuroimaging of emotion:
752 empirical effects of proportional global signal scaling in fMRI data analysis.
753 *NeuroImage, 25*(2), 520-526. doi:10.1016/j.neuroimage.2004.12.011
754 Kentridge, R. W., Thomson, R., & Heywood, C. A. (2012). Glossiness perception can be
755 mediated independently of cortical processing of colour or texture. *Cortex, 48*(9),
756 1244-1246. doi:10.1016/j.cortex.2012.01.011
757 Kerrigan, I. S., & Adams, W. J. (2013). Highlights, disparity, and perceived gloss with
758 convex and concave surfaces. *Journal of Vision, 13*(1). doi:10.1167/13.1.9
759 Kim, J., & Anderson, B. L. (2010). Image statistics and the perception of surface gloss and
760 lightness. *Journal of Vision, 10*(9). doi:10.1167/10.9.3
761 Kim, J., Marlow, P., & Anderson, B. L. (2011). The perception of gloss depends on highlight
762 congruence with surface shading. *Journal of Vision, 11*(9). doi:10.1167/11.9.4
763 Kim, J., Marlow, P. J., & Anderson, B. L. (2012). The dark side of gloss. *Nature*
764 *Neuroscience, 15*(11), 1590-1595. doi:10.1038/nn.3221
765 Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain
766 mapping. *Proceedings of the National Academy of Sciences of the United States of*
767 *America, 103*(10), 3863-3868. doi:10.1073/pnas.0600244103
768 Landy, M. S. (2007). Visual perception: A gloss on surface properties. *Nature, 447*(7141),
769 158-159. doi:10.1038/nature05714
770 Marlow, P., & Anderson, B. L. (2013). Generative constraints on image cues for perceived
771 gloss. *Journal of Vision, 13*(14). doi:10.1167/13.14.2
772 Marlow, P., Kim, J., & Anderson, B. L. (2011). The role of brightness and orientation
773 congruence in the perception of surface gloss. *Journal of Vision, 11*(9).
774 doi:10.1167/11.9.16

775 Marlow, P., Kim, J., & Anderson, Barton L. (2012). The Perception and Misperception of
776 Specular Surface Reflectance. *Current Biology, 22*(20), 1909-1913.
777 doi:10.1016/j.cub.2012.08.009
778 Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the
779 perception of surface qualities. *Nature, 447*(7141), 206-209. doi:10.1038/nature05724
780 Murphy, A. P., Ban, H., & Welchman, A. E. (2013). Integration of texture and disparity cues
781 to surface slant in dorsal visual cortex. *Journal of Neurophysiology, 110*(1), 190-203.
782 doi:10.1152/jn.01055.2012
783 Muryy, A. A., Fleming, R. W., & Welchman, A. E. (2012). Binocular cues for glossiness.
784 *Journal of Vision, 12*(9), 869. doi:10.1167/12.9.869
785 Muryy, A. A., Fleming, R. W., & Welchman, A. E. (2014). Key characteristics of specular
786 stereo. *Journal of Vision, 14*(14), 1-26. doi:10.1167/14.14.14
787 Muryy, A. A., Welchman, A. E., Blake, A., & Fleming, R. W. (2013). Specular reflections
788 and the estimation of shape from binocular disparity. *Proceedings of the National*
789 *Academy of Sciences, 110*(6), 2413-2418. doi:10.1073/pnas.1212417110
790 Neri, P., Bridge, H., & Heeger, D. J. (2004). Stereoscopic Processing of Absolute and
791 Relative Disparity in Human Visual Cortex. *Journal of Neurophysiology, 92*(3), 1880-
792 1891. doi:10.1152/jn.01042.2003
793 Nishio, A., Goda, N., & Komatsu, H. (2012). Neural Selectivity and Representation of Gloss
794 in the Monkey Inferior Temporal Cortex. *The Journal of Neuroscience, 32*(31),
795 10780-10793. doi:10.1523/jneurosci.1095-12.2012
796 Nishio, A., Shimokawa, T., Goda, N., & Komatsu, H. (2014). Perceptual Gloss Parameters
797 Are Encoded by Population Responses in the Monkey Inferior Temporal Cortex. *The*
798 *Journal of Neuroscience, 34*(33), 11143-11151. doi:10.1523/jneurosci.1451-14.2014

799 Obein, G., Knoblauch, K., & Viénot, F. (2004). Difference scaling of gloss: Nonlinearity,
800 binocularity, and constancy. *Journal of Vision, 4*(9). doi:10.1167/4.9.4
801 Okazawa, G., Goda, N., & Komatsu, H. (2012). Selective responses to specular surfaces in
802 the macaque visual cortex revealed by fMRI. *NeuroImage, 63*(3), 1321-1333.
803 doi:10.1016/j.neuroimage.2012.07.052
804 Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world
805 illumination. *Journal of Vision, 10*(9). doi:10.1167/10.9.5
806 Orban, G. A., Claeys, K., Nelissen, K., Smans, R., Sunaert, S., Todd, J. T., . . . Vanduffel, W.
807 (2006). Mapping the parietal cortex of human and non-human primates.
808 *Neuropsychologia, 44*(13), 2647-2667. doi:10.1016/j.neuropsychologia.2005.11.001
809 Orban, G. A., Fize, D., Peuskens, H., Denys, K., Nelissen, K., Sunaert, S., . . . Vanduffel, W.
810 (2003). Similarities and differences in motion processing between the human and
811 macaque brain: evidence from fMRI. *Neuropsychologia, 41*(13), 1757-1768.
812 doi:dx.doi.org/10.1016/S0028-3932(03)00177-5
813 Orban, G. A., Sunaert, S., Todd, J. T., Van Hecke, P., & Marchal, G. (1999). Human Cortical
814 Regions Involved in Extracting Depth from Motion. *Neuron, 24*(4), 929-940.
815 doi:10.1016/S0896-6273(00)81040-5
816 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming
817 numbers into movies. *Spatial Vision, 10*(4), 437-442. doi:10.1163/156856897X00366
818 Preston, T. J., Kourtzi, Z., & Welchman, A. E. (2009). Adaptive Estimation of Three-
819 Dimensional Structure in the Human Brain. *The Journal of Neuroscience, 29*(6),
820 1688-1698. doi:10.1523/jneurosci.5021-08.2009
821 Sakano, Y., & Ando, H. (2010). Effects of head motion and stereo viewing on perceived
822 glossiness. *Journal of Vision, 10*(9). doi:10.1167/10.9.15

Sharan, L., Li, Y., Motoyoshi, I., Nishida, S., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of the Optical Society of America, 25*(4), 846-865. Retrieved from http://josaa.osa.org/abstract.cfm?URI=josaa-25-4-846

Sun, H.-C., Ban, H., Di Luca, M., & Welchman, A. E. (2015). fMRI evidence for areas that process surface gloss in the human visual cortex. *Vision Research, 109*, 149-157. doi:10.1016/j.visres.2014.11.012

Sun, H.-C., Welchman, A. E., Chang, D. H. F., & Di Luca, M. (2016). Look but don't touch: Visual cues to surface structure drive somatosensory cortex. *NeuroImage*. doi:10.1016/j.neuroimage.2015.12.054

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., & Orban, G. A. (2002). Extracting 3D from Motion: Differences in Human and Monkey Intraparietal Cortex. *Science, 298*(5592), 413-415. doi:10.1126/science.1073574

Wada, A., Sakano, Y., & Ando, H. (2014). Human cortical areas involved in perception of surface glossiness. *NeuroImage, 98*(0), 243-257. doi:10.1016/j.neuroimage.2014.05.001

Wendt, G., Faul, F., Ekroll, V., & Mausfeld, R. (2010). Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision, 10*(9). doi:10.1167/10.9.7

Wendt, G., Faul, F., & Mausfeld, R. (2008). Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision, 8*(1). doi:10.1167/8.1.14

Figure 1

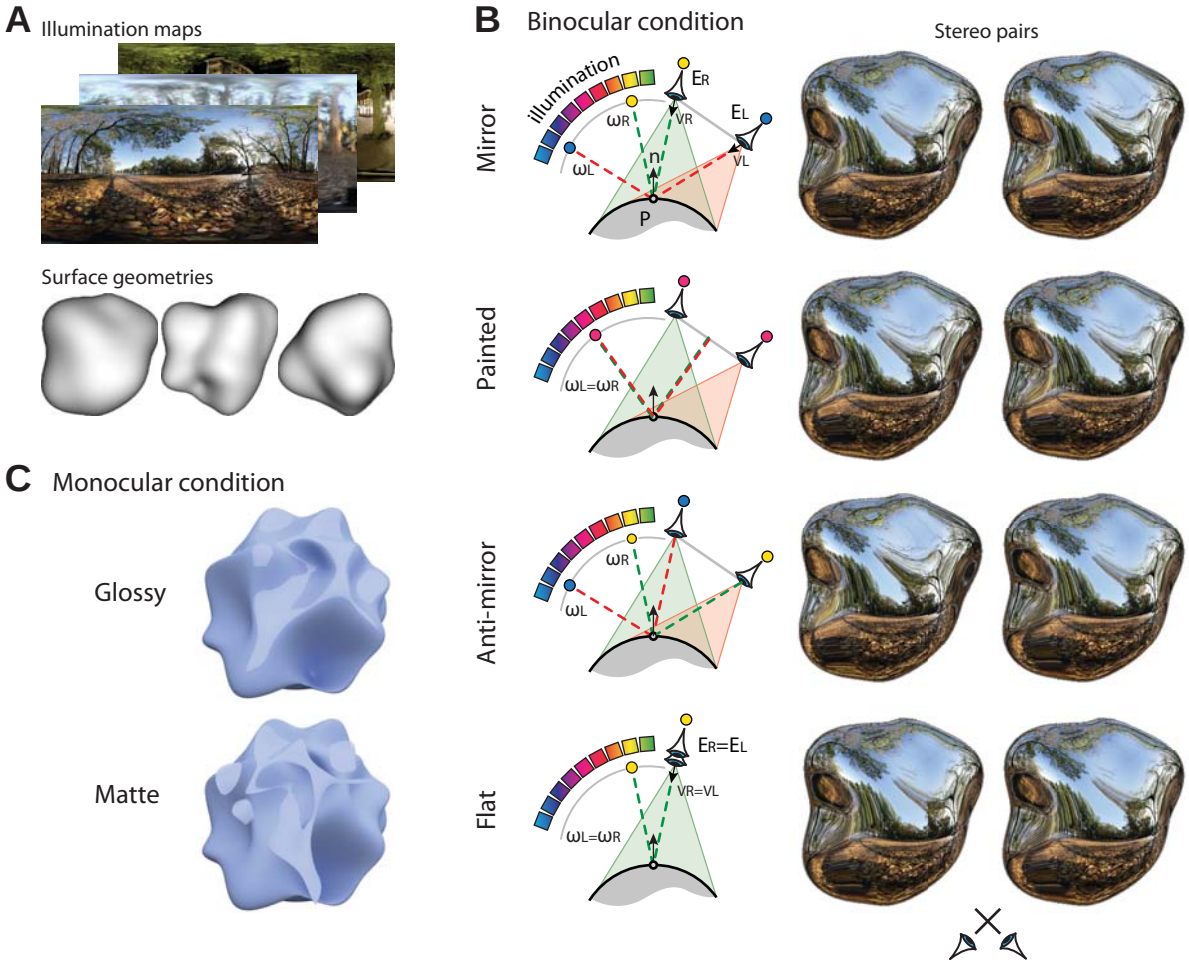

A
Illumination maps
Surface geometries
B
Binocular condition
Stereo pairs
Mirror
Painted
Anti-mirror
Flat
illumination
ωR
ωL
ER
VR
EL
VL
n
P
ωL=ωR
ER=EL
VR=VL
C
Monocular condition
Glossy
Matte

Figure 2

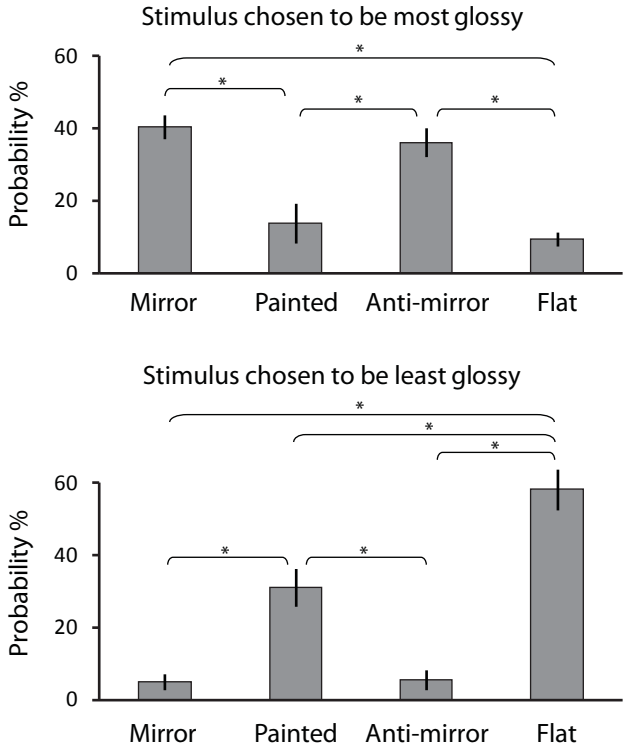

Stimulus chosen to be most glossy
Probability %
60
40
20
0
*
*
*
*
Mirror
Painted
Anti-mirror
Flat
Stimulus chosen to be least glossy
Probability %
60
40
20
0
*
*
*
*
*
Mirror
Painted
Anti-mirror
Flat

Figure 3

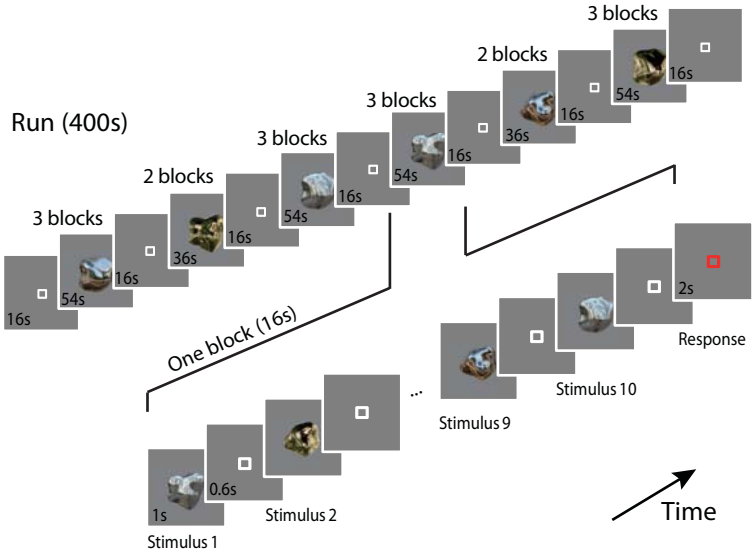

Run (400s)
3 blocks
16s
54s
2 blocks
16s
36s
3 blocks
16s
54s
3 blocks
16s
54s
2 blocks
16s
36s
3 blocks
16s
54s
One block (16s)
1s
0.6s
Stimulus 1
Stimulus 2
...
Stimulus 9
Stimulus 10
2s
Response
Time

Figure 4

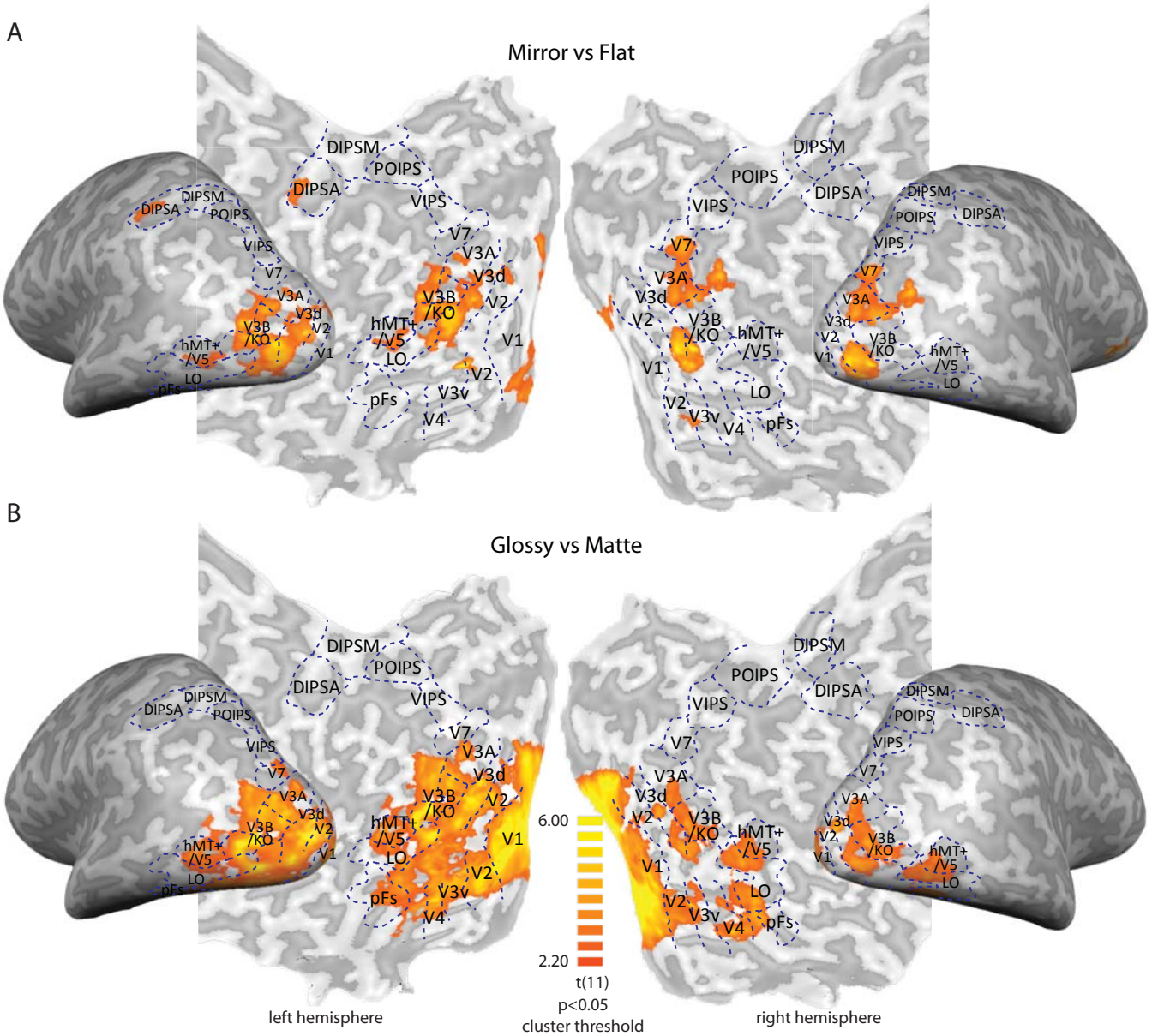

A
Mirror vs Flat
B
Glossy vs Matte
DIPSM
POIPS
DIPSA
VIPS
V7
V3A
V3d
V3B
/KO
V2
V1
hMT+
/V5
LO
pFs
V3v
V4
6.00
2.20
t(11)
p<0.05
cluster threshold
left hemisphere
right hemisphere

Figure 5

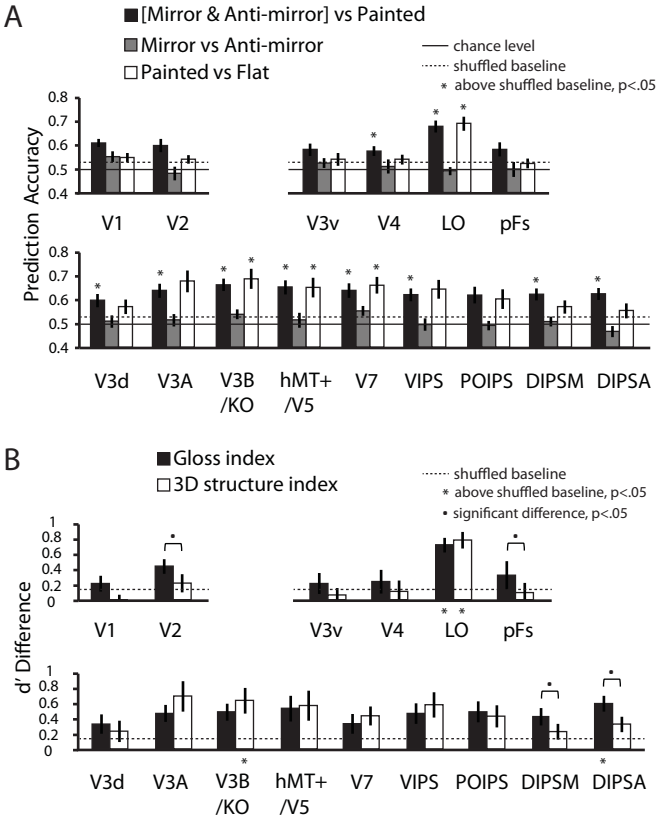

A
[Mirror & Anti-mirror] vs Painted
Mirror vs Anti-mirror
Painted vs Flat
chance level
shuffled baseline
* above shuffled baseline, p<.05
Prediction Accuracy
0.8
0.7
0.6
0.5
0.4
V1
V2
V3v
V4
LO
pFs
V3d
V3A
V3B /KO
hMT+ /V5
V7
VIPS
POIPS
DIPSM
DIPSA
B
Gloss index
3D structure index
shuffled baseline
* above shuffled baseline, p<.05
• significant difference, p<.05
d' Difference
1
0.8
0.6
0.4
0.2
0
V1
V2
V3v
V4
LO
pFs
V3d
V3A
V3B /KO
hMT+ /V5
V7
VIPS
POIPS
DIPSM
DIPSA

Figure 6

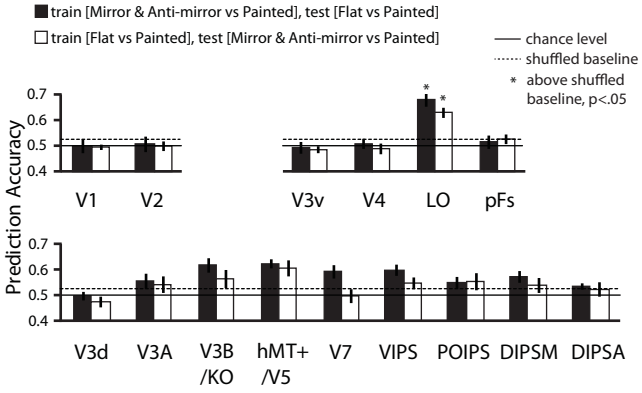

train [Mirror & Anti-mirror vs Painted], test [Flat vs Painted]
train [Flat vs Painted], test [Mirror & Anti-mirror vs Painted]
chance level
shuffled baseline
* above shuffled baseline, p<.05
Prediction Accuracy
0.7
0.6
0.5
0.4
V1
V2
V3v
V4
LO
pFs
V3d
V3A
V3B /KO
hMT+ /V5
V7
VIPS
POIPS
DIPSM
DIPSA

Figure 7

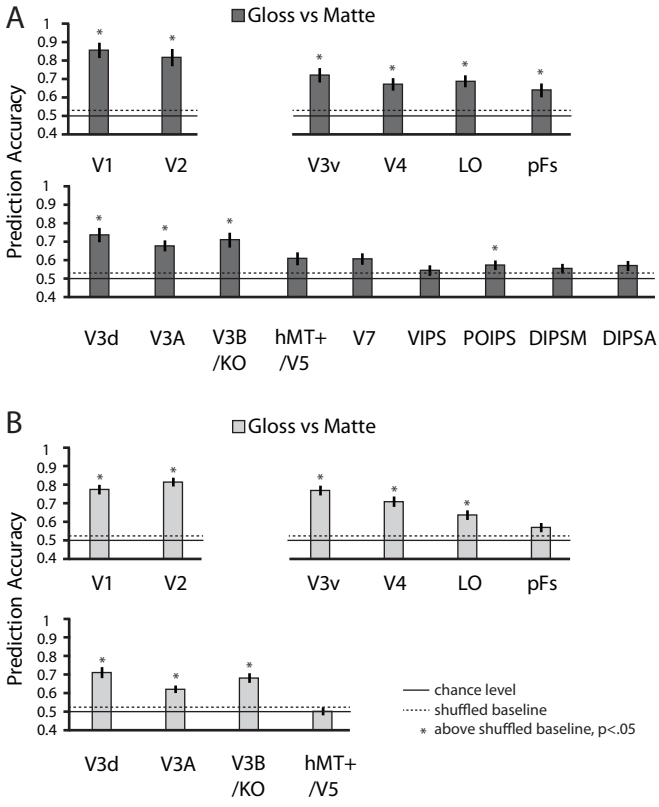

A
Gloss vs Matte
Prediction Accuracy
1
0.9
0.8
0.7
0.6
0.5
0.4
V1
V2
V3v
V4
LO
pFs
V3d
V3A
V3B
/KO
hMT+
/V5
V7
VIPS
POIPS
DIPSM
DIPSA
B
Gloss vs Matte
Prediction Accuracy
chance level
shuffled baseline
* above shuffled baseline, p<.05

Figure 8

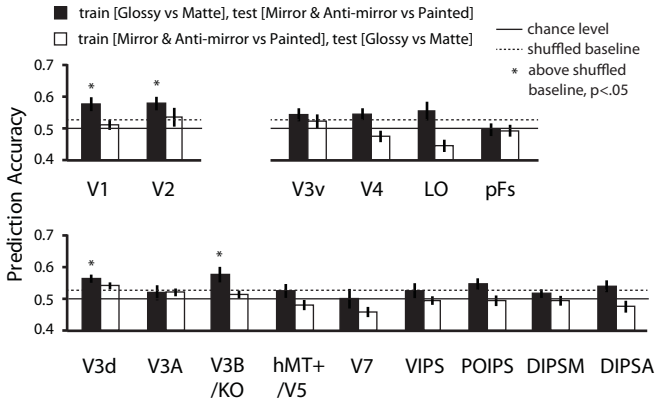
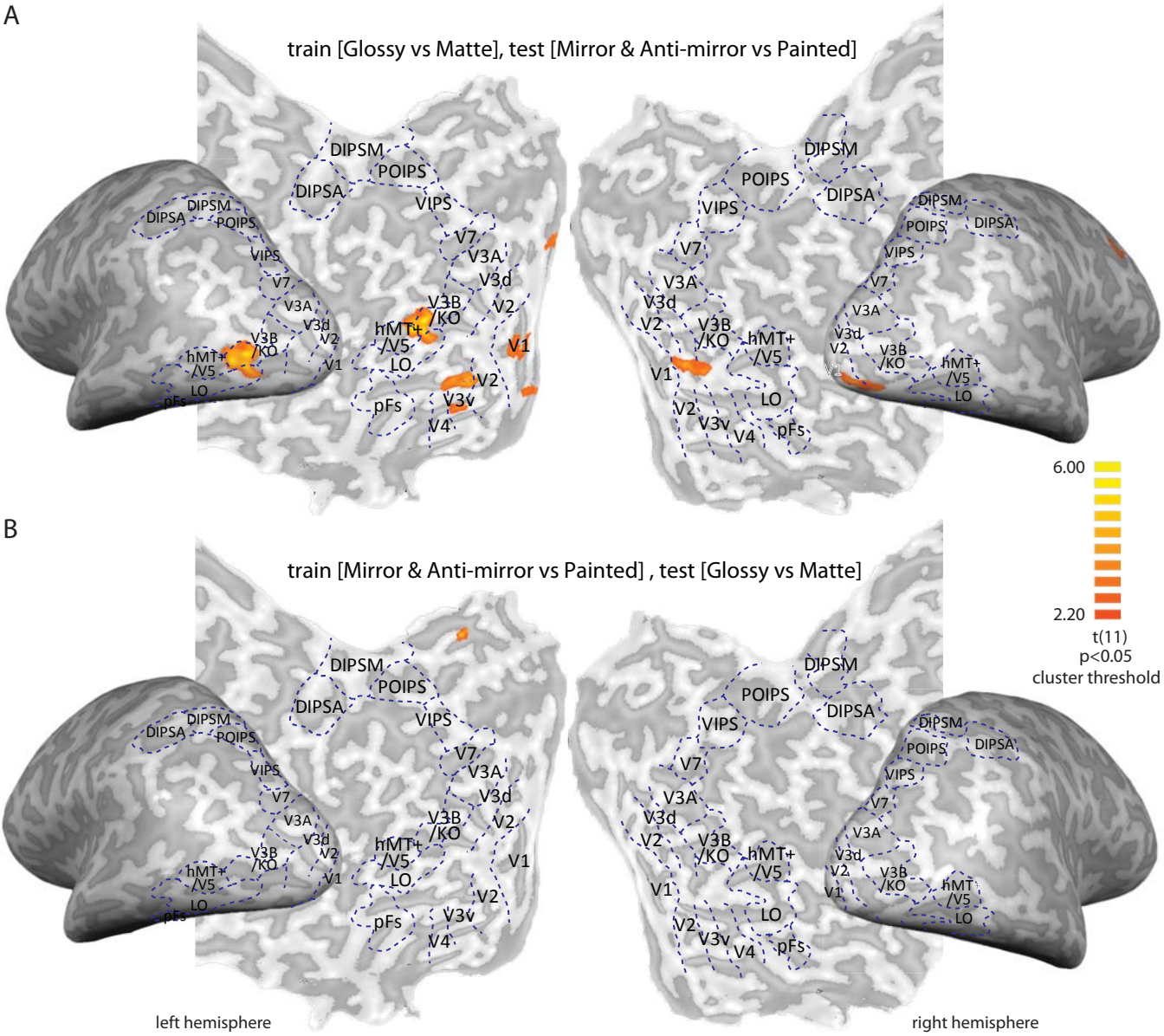

train [Glossy vs Matte], test [Mirror & Anti-mirror vs Painted]
train [Mirror & Anti-mirror vs Painted], test [Glossy vs Matte]
chance level
shuffled baseline
* above shuffled baseline, p<.05
Prediction Accuracy
0.7
0.6
0.5
0.4
V1
V2
V3v
V4
LO
pFs
V3d
V3A
V3B /KO
hMT+ /V5
V7
VIPS
POIPS
DIPSM
DIPSA

Figure 9


A
train [Glossy vs Matte], test [Mirror & Anti-mirror vs Painted]
B
train [Mirror & Anti-mirror vs Painted] , test [Glossy vs Matte]
DIPSM
POIPS
DIPSA
VIPS
V7
V3A
V3d
V2
V1
V3B
/KO
hMT+
/V5
LO
pFs
V3v
V4
6.00
2.20
t(11)
p<0.05
cluster threshold
left hemisphere
right hemisphere