OXFORD

Cultural specialization of visual cortex

John C. Ksander, Laura E. Paige, Hunter A. Johndro, and Angela H. Gutchess

Department of Psychology, Brandeis University, Waltham, MA 02453, USA

Correspondence should be addressed to John Ksander, Department of Psychology, Brandeis University, 415 South Street, MS 062, Waltham, MA 02453, USA. E-mail: jksander@brandeis.edu.

Abstract

A growing body of evidence suggests culture influences how individuals perceive the world around them. This study investigates whether these cultural differences extend to a simple object viewing task and visual cortex by examining voxel pattern representations with multi-voxel pattern analysis (MVPA). During functional magnetic resonance imaging scanning, 20 East Asian and 20 American participants viewed photos of everyday items, equated for familiarity and conceptual agreement across cultures. Whole brain searchlight mapping with non-parametric statistical evaluation tested whether these stimuli evoked multi-voxel patterns that were distinct between cultural groups. We found that participants' cultural identities were successfully predicted from stimuli representations in visual cortex Brodmann areas 18 and 19. This result demonstrates culturally specialized visual cortex during a basic perceptual task ubiquitous to everyday life.

Key words: fMRI; MVPA; cross-cultural; perception

Cultural specialization of visual cortex

Recent findings in psychology and neuroscience demonstrate that culture substantially impacts how individuals perceive the world around them. This evidence comes from experiments testing explicitly cultural behavior, such as reading in one's native *vs* non-native language (e.g. Baker *et al*., 2007), as well as experiments testing culturally-nonspecific behavior, such as simple visuospatial tasks (e.g. Hedden *et al*., 2008). Multiple frameworks, the neuronal recycling hypothesis and theories of cultural differences in information processing, account for how culture might shape perception.

The Dehaene and Cohen (2007) neuronal recycling theory hypothesizes that cortical specialization enables culturally acquired behavior. Cultural specialization is, in turn, strongly constrained by that cortex's original functionality. For example, reading a written language would specialize the cortex performing the prototypical computations required for that ability. Reading is then both enabled by this specialized cortex and constrained by that cortex's functional origins. Evidence for visual word form area sensitivity to written language (Cohen and Dehaene, 2004) and language-specific functionality (Paulesu *et al*., 2000) supports this account. The neuronal recycling

hypothesis mainly concerns cultural inventions, which are optional, evolutionarily recent behaviors acquired through learning (e.g. written language, arithmetic and some tool usage).

Another prevalent theory proposes that culturally specialized perception reflects information processing biases. For example, culture can influence estimations of line lengths within different visual contexts (Kitayama *et al*., 2003; Hedden *et al*., 2008), focal object processing (Gutchess *et al*., 2006), and change-detection ability (Boduroglu *et al*., 2009). Some of these biases are discussed in the literature as differences in 'analytic' *vs* 'holistic' approaches. More specifically, individuals from Western cultures prioritize feature-based information, whereas individuals from Eastern cultures prioritize contextually based information (e.g. Masuda and Nisbett, 2001; Nisbett *et al*., 2001; Masuda *et al*., 2008; Rule *et al*., 2013). For example, when viewing a photograph depicting a group of people, a Westerner may preferentially attend to the photograph's visual details, such as the groups' clothes, or hair color. An Easterner may instead preferentially attend to contextual information, such as relationships amongst individuals in the group, or the depicted group's location. These biases have been linked to

corresponding differences in visual search (Wang *et al.*, 2012) and recognition memory performance [Millar *et al.*, 2013; see Paige *et al.*, 2017a for a corresponding functional magnetic resonance imaging (fMRI) study].

Previous behavioral and fMRI studies have examined culturally biased information processing using simple visuospatial tasks and judgments of complex scenes. The behavioral findings largely support culturally specialized perception, in that Easterners prioritize contextual visual information while Westerners prioritize feature-based information, as reviewed by Gutchess and Indeck (2009). However, it is important to note that some behavioral experiments have not found perceptual differences between culture groups, and the reported behavioral findings do not unanimously support culturally specialized perception (e.g. Zhou *et al.*, 2008; Evans *et al.*, 2009). Unlike the behavioral literature, fMRI results uniformly show cultural differences in blood-oxygen-level dependent (BOLD) activity for perceptual tasks, based on two comprehensive reviews of cross-cultural fMRI studies (Rule *et al.*, 2013; Han and Ma, 2014).

Therefore, the rich neural data afforded by neuroimaging methods may be advantageous for studying culturally specialized perception. Kitayama and Uskul (2011) argue that culture-specific cognition has a strong neural basis and consequently, neuroimaging experiments show greater sensitivity compared with behavioral studies. Modern neuroimaging techniques would then seem particularly suitable for evaluating whether perception involves culturally specialized cortex in experiments with conflicting behavioral results, such as simple visual tasks without explicitly culturally laden stimuli.

This study investigates whether perception involves culturally specialized cortex during a simple object viewing task with multi-voxel pattern analysis (MVPA). Previous fMRI studies investigating culture and visual perception have exclusively utilized univariate fMRI analyses [see reviews by Chiao (2009), Han and Ma (2014) and Rule *et al.* (2013)], which only reveal where average brain activity differs, typically after spatial smoothing. This approach is likely suboptimal for this study, as regional BOLD activity may not reflect the fine-grained cortical specialization that is likely necessary to produce perceptual differences between cultures. How people perceive the world differently is also arguably multivariate by nature, and therefore poorly described by univarate measurements [see review by Charest and Kriegeskorte (2015)]. Unlike univariate approaches, MVPA methodology assumes information in the brain is represented by a distributed neural code (i.e. fine-grained patterns of voxel activity across cortex) (Norman *et al.*, 2006). This offers much greater analytical sensitivity, particularly in visual perception experiments (Haxby *et al.*, 2001; Serences and Boynton, 2007). Therefore, MVPA provides a more suitable approach for this study, despite its underutilization to date in cultural neuroscience. The authors are aware of only one other cross-cultural study employing this approach, in which Raizada *et al.* (2010) demonstrated that voxel patterns can predict how well Japanese and English speakers are able to discriminate between syllables specific to the English language. This study also concluded that univariate analyses insufficiently characterized the distinctions between cross-cultural representations of auditory stimuli, as mean BOLD signals did not show this effect.

This study investigates perception, employing a simple object viewing task with stimuli that are not specific to any particular culture. East Asian and American participants viewed pictures of common objects, which were equated for cultural familiarity in a previous study, and then further checked for cultural equivalence in this study. If these stimuli evoke multi-voxel patterns that are distinct between cultural groups, this will provide compelling evidence that perception involves culturally specialized cortex.

Materials and methods

Dataset

This study re-analyzes a dataset (Paige *et al.*, 2017b) originally collected to investigate cultural specificity in memory. The present analyses evaluate perception, distinct from memory outcomes.

Participants

Twenty East Asian (10 female) and 20 American (17 female) participants were recruited for this study. East Asian participants originated from China (including Hong Kong and Taiwan), Japan, Korea, Malaysia, Thailand or Vietnam and had lived in the United States for < 5 years ($M = 2.6$, s.d. = 1.3, range = 0.5–4.5) prior to the experiment. All American participants were United States natives, native English speakers (learned before 5 years of age), and had not resided outside the United States for >5 years.

All participants were between the ages of 18 and 35 (Americans $M = 22.5$, s.d. = 2.4; East Asians $M = 24.3$, s.d. = 3.8). All participants were right handed, had normal (or corrected-to-normal) vision and hearing and were screened for head trauma (loss of consciousness for >10 min), emotional, psychiatric or learning disorders, and other contraindications for scanning. Participants provided written informed consent for the protocol approved by the Brandeis University IRB.

Stimuli

The image stimuli were developed by Kensinger *et al.* (2007), and later used by Millar *et al.* (2013) for testing cross-cultural differences in memory performance. Millar *et al.* (2013) piloted the stimuli with American and East Asian participants to match cultural familiarity and conceptual agreement. Stimuli were images of common objects against white backgrounds (examples in Figure 1). In total 216 images were used for encoding, with two different exemplars sharing the same verbal label (e.g. apple) employed across participants.

Picture viewing procedure

Each participant viewed 108 pictures during fMRI scanning, divided into two runs of 54 images. The selection of individual images was counterbalanced across four selection lists, determining which stimuli appeared in each run for each participant, and equated across cultures. The stimuli order within each run was randomized for each participant.

At each trial, a prompt, presented in English, was shown for 2 s: 'Please indicate whether you would approach/avoid/stay'. This was followed by an image for 500 ms and a fixation cross during the interstimulus interval from 3500 to 11 500 ms, with jitter determined by Optseq (Dale, 1999). When the image appeared, participants made their decision and responded with a button press. Before scanning, participants practiced the task until competent.

The 'approach/avoid/stay' trial prompts were intended to ensure participants paid attention and encoded items throughout the experiment, as the original experiment included a surprise memory test. The prompt responses and retrieval

Fig. 1. Four example stimuli.

performance are not of interest in this study, and will not be discussed further [see Paige *et al.* (2017a) for further treatment of this issue].

Stimulus familiarity

Approximately 10 days after encoding, participants completed an online survey in which they named and rated their familiarity with each object in the stimulus set (1–5 Likert rating scale). This survey assessed whether participants across cultures reported equal familiarity with the stimuli. Even though these stimuli were previously equated for cultural familiarity in Millar *et al.* (2013), this check is important for interpreting the current results so that differences between culture groups cannot simply be attributed to differences in stimulus familiarity.

fMRI data acquisition

Images were acquired using a Siemens Trio 3T whole-body scanner. The two encoding runs utilized a 32-channel head coil and simultaneous multi-slice scanning, for which 69 slices 2.0-mm thick were acquired with an echo-planar image sequence (TR = 2000 ms, TE = 30 ms, FOV = 216 mm, and flip angle = 80°). Stimuli presentations were timelocked to TR onsets. High-resolution anatomical images were acquired using multiplanar rapidly acquired gradient echo sequence.

fMRI pre-processing

Pre-processing was implemented with SPM12 (Wellcome Department of Cognition Neurology, London, UK). Images were slice-time corrected, realigned for motion correction, co-registered to anatomical data, and normalized to the Montreal Neurological Institute (MNI) template space. No spatial smoothing was performed after normalization; subsequent analyses were implemented with in-house Matlab scripts.

A gray matter tissue probability map was applied to the unsmoothed data, and only voxels with at least 10% gray matter probability were retained. This tissue probability map was constructed with the IXI dataset (http://www.brain-development.org/), the same tissue probability map used in SPM segmentation and spatial normalization routines.

Hemodynamic response modeling

The hemodynamic response (HDR) for each stimulus was modeled using the least squares-separate (LS-S) method described in Mumford *et al.* (2012). Event-related fMRI designs with short interstimulus intervals, like this study, pose a general challenge for MVPA. Trial HDRs overlap in these designs, so activity during the HDR of one trial will often contain signal from a previous

trial as well. The LS-S technique addresses this problem by deconvolving BOLD activity for individual trials from an experiment's fMRI timecourse. This allows the estimation of voxel response patterns specific to individual stimuli.

LS-S HDR modeling was performed separately for each subject and within individual runs. Modeling each run separately maintains the independence between training and testing datasets necessary for cross-validation in the subsequent analyses (Kriegeskorte *et al.*, 2009). Therefore, the following modeling procedure was applied individually to each of the 80 runs in the dataset (i.e. 40 subjects × 2 runs, every run modeled separately). Thus, the following procedure was applied to a single run from a single subject, and iteratively repeated across the whole dataset.

First, each voxel's timecourse was normalized to zero mean and unit variance (i.e. Z scored), then a highpass temporal filtering was applied (Gaussian-weighted least-squares line, 128 s cut-off). The LS-S design matrix consisted of an intercept and two regressors. Following the procedure in Mumford *et al.* (2012), the first regressor modeled the trial of interest HDR, and the second modeled all other trial HDRs. The 128 s cut-off highpass filter was also applied to these regressors prior to HDR estimation. For every trial, HDR voxel betas were estimated, and then *t*-values were calculated by dividing the beta estimates by their standard errors. Transforming response pattern betas into *t*-values is a noise normalization technique shown to limit the influence of noisy voxels on classification analyses (Misaki *et al.*, 2010). The estimated patterns for trials which began less than 3 TRs before the end of a run were discarded for insufficient data. Thus, the response pattern to each stimulus was measured by the voxel *t*-values calculated from LS-S modeling.

Modeling each trial's HDR pattern in this way served two purposes. First, stimulus-specific response patterns were obtained from the event-related timecourse. Second, this procedure should separate the activity related to the image stimuli from the activity related to the trial prompts that preceded every trial. As it has been demonstrated that reading in one's native *vs* non-native language produces differences in mean BOLD activity (Baker *et al.*, 2007), the sensitivity of MVPA combined with this study's overlapping trial HDRs presents the possibility that our results could be confounded with neural responses from reading the English trial prompts. Although participants likely habituated to the prompt (repeated throughout the study and practice), the possibility of differences resulting from native language was addressed analytically. The LS-S modeling performed in this study deconvolves HDRs timelocked to the image stimulus onsets, which suppresses prestimulus 'bleed through' activity. This model further accounts for activity common between trials by including an intercept term. Therefore, HDR estimates would not include prestimulus bleed

through activity common to all trials, such as leftover activity from a common prompt.

Multi-voxel pattern analysis

Pre-processing. Two further pre-processing steps were taken for these analyses; the first step removed unsuitable fMRI data, and the second step improved the suitable data's signal-to-noise ratio. Only voxels valid in all subjects are appropriate for this study's classification design (see 'Analysis design' section), so voxels with signal dropout in any subject at any timepoint during the experiment were considered invalid and removed (see Supplementary Material).

Finally, run-wise temporal compression was performed on the valid voxels. This procedure averages the trial HDR estimates within each run, producing two voxel response patterns per subject (one per run). Temporal compression improves signal-to-noise ratio without spatial smoothing by averaging over noisy trial-to-trial pattern variance [Mourao-Miranda *et al.* (2006), see Supplementary Material]. These temporally compressed *t*-value patterns were the input data for all analyses.

Analysis design. The preprocessed data was submitted to a leave-one-subject-out (LOSO) classification design. In this scheme, a classifier is trained on every subject except one left-out subject, then tested on the left-out subject. Model cross-validation is achieved by repeating this process iteratively until every subject has been left-out once (*k*-folds = 40).

The classifier was trained with fMRI activation patterns and corresponding participant group labels (i.e. East Asian or American) in the training dataset, and predicted the unseen participant labels of response patterns in the testing dataset. In other words, the classifier was trained to distinguish East Asian from American response patterns, and then predicted whether the left-out subject's response patterns were from an East Asian or an American participant. Classification accuracy was measured by the percentage of correctly predicted labels across all cross-validation folds, or, equivalently, the average LOSO classification accuracy.

Searchlight mapping. Searchlight MVPA approaches analyze response patterns contained within smaller regions of interest (ROIs) iteratively across the whole brain (Kriegeskorte *et al.*, 2006). This allows the localization of culturally specialized cortex without a priori ROIs in this study. Spherical searchlights 3 voxels (6 mm) in diameter were used, each with a volume of 19 total voxels. The average LOSO classification accuracy for each searchlight was mapped to its center voxel. Importantly, searchlights were only analyzed in locations where all voxels within the searchlight were valid for analysis. Searchlights containing any voxels excluded during preprocessing (invalid for LOSO cross-validation or <10% gray matter probability) were not analyzed. These parameters created 145 938 total searchlights.

Classifier. Searchlights were analyzed with a Gaussian Naive Bayes (GNB) classifier according to the previously described LOSO design. The GNB classifier was implemented as detailed in Raizada and Lee (2013), with a class prior term for the slightly unbalanced LOSO training set:

$$\hat{y} = \underset{k \in \{1, \ldots K\}}{\operatorname{argmax}} \ \log p(C_k) + \sum_{i=1}^{n} \log p(x_i | C_k)$$

The classifier predicts one of *K* classes, C_k, based on the largest class prior, $p(C_k)$, and voxel membership, $p(x_i | vC_k)$, joint probability. Class prior probability was defined as $p(C_k) = \frac{N_k}{N}$ where *N* is the total number of training observations (i.e. patterns) and N_k is the number of training observations belonging to class C_k. The equation for voxel class likelihood (i.e. the probability an individual searchlight voxel, x_i, belongs to a given class) is defined as:

$$p(x_i | vC_k) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x_i - \mu_k)^2 / 2\sigma^2_k}$$

The major advantage of searchlight MVPA with Gaussian Naive Bayes classification is that GNB's 'naiveté' can be exploited during cross-validation for dramatically superior computation time compared with other algorithms. This fast GNB implementation was first reported by Pereira and Botvinick (2011), then expanded on by Raizada and Lee (2013). The GNB model (naively) assumes that class features are independent of one another (e.g. the correlations between response pattern voxels do not matter). Therefore, the GNB model can be trained on every voxel in the whole brain all at once, whereas other algorithms must be trained on each searchlight individually in order to model the feature relationships in different voxel groupings. The number of CV folds compounds this advantage further, as each fold requires model retraining. Importantly, this fast GNB implementation makes searchlight permutation testing computationally tractable.

Searchlight mapping statistical inference. A non-parametric permutation testing procedure was used to evaluate the statistical significance of searchlight classification accuracies. This approach is particularly warranted as recent publications have raised concerns over parametric significance testing in fMRI (Eklund *et al.*, 2012, 2016; Stelzer *et al.*, 2013). Therefore, the non-parametric cluster significance test developed by Stelzer *et al.* (2013) was adapted for the current between-subjects MVPA design. Precise details of how we implemented this method are in the Supplementary Material, but this adaptation only deviated from the original Stelzer *et al.* (2013) procedure in one major regard. Specifically, group accuracy maps were not bootstrapped from single subject permutation maps. This is simply because the current analysis is between-subjects and thus does not involve single subject searchlight maps or group average accuracy maps. This required more computationally expensive permutation testing, although the permuted searchlight maps in this study were analyzed in the same fashion as the bootstrapped group accuracy maps in Stelzer *et al.* (2013).

Computation time

The searchlight mapping analysis was implemented with the Matlab parallel computing toolbox for running on a single server node (Brandeis University High Performance Computing Cluster) with 512GB RAM and 4× Intel Xeon CPU E5-4620 v2 @ 2.60 GHz, each with 8 physical cores (32 cores total). The total computation time was ∼15 days for this analysis. Even with a parallelized fast-GNB implementation and powerful computing resources, searchlight permutation testing accounted for >14 days of that run time. This computational limitation is why the fMRI analyses enjoyed fewer permutations during statistical evaluation than the behavioral analysis (see 'Familiarity Check' section).
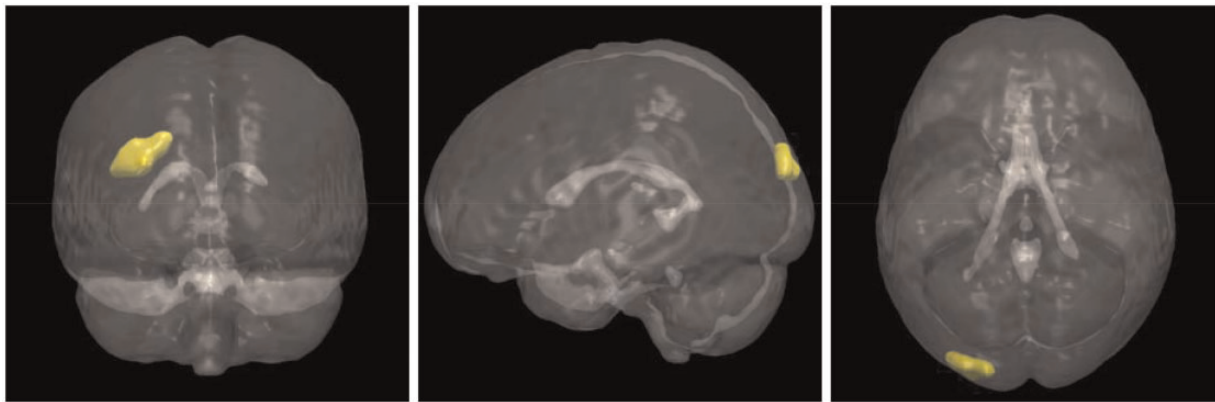
Fig. 2. Searchlight mapping results. All voxels in the 41 searchlight cluster spanning BA 18 and 19 are shown in yellow. Glass brain rendered via Madan (2015).

Results

Familiarity check

The post-task questionnaire revealed no differences in stimulus familiarity (P = 0.380) between East Asians (M = 4.38, s.d. = 0.20) and Americans (M = 4.30, s.d. = 0.31), tested via permutation testing of mean differences with 50 000 permutations. This finding replicates the Millar *et al.* (2013) analyses, which also determined these stimuli were equally familiar across cultures.

Searchlight mapping

The whole brain searchlight mapping analysis revealed one significant searchlight cluster after non-parametric cluster size control (FDR $q = 0.05$, $p_{corrected} = 0.03$, $q_{threshold} = 0.0007$, $p_{uncorrected} = 0.0004$). This cluster comprised 41 contiguous searchlights spanning Brodmann areas 18 and 19 in visual cortex (see Figure 2). The MNI coordinates and Brodmann areas of each searchlight center in this cluster are listed in the Supplementary Material.

ROI follow-up analyses

In order to determine whether culturally distinct neural codes are distributed across the visual cortex identified by searchlight mapping, follow-up ROI analyses were performed as prescribed by Etzel *et al.* (2013). In these analyses, an ROI was constructed from the significant searchlight cluster[1] and tested. This analysis is deliberately circular (Kriegeskorte *et al.*, 2009), as discussed by Etzel *et al.* (2013). The purpose is to test whether the cortex identified in searchlight mapping is informative as a whole, rather than just an agglomeration of smaller informative searchlights. However, this analysis also afforded the opportunity to test three important qualities of our result. First, whether the effect generalized across different classifier algorithms, as testing such generalizability with searchlight mapping was computationally intractable. Second, whether this result truly reflected fine-grained neural coding, rather than differences in mean activity. Third, how differences in anatomical alignment or gender ratios may have influenced this result. These empirical tests were further characterized by visualizing the ROI data for illustrative purposes (see 't-Distributed stochastic neighbor embedding visualization' section).

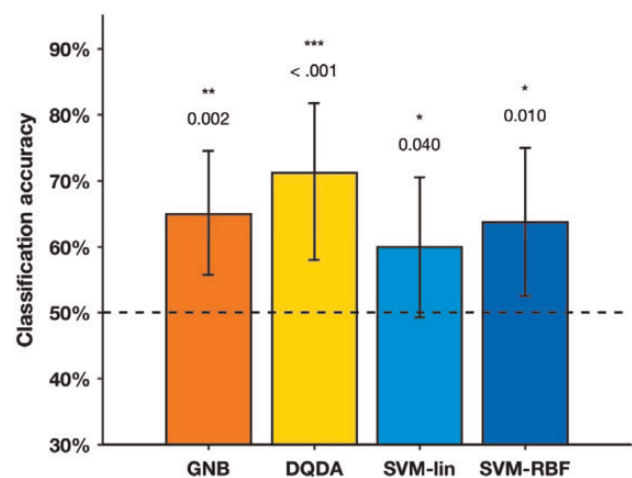1 All searchlight voxels were included, not just the center voxels.

Fig. 3. ROI follow-up classification accuracies. The mean empirical chance accuracies were 49.29, 49.5, 49.47 and 48.82% for GNB, DQDA, SVM-lin and SVM-RBF, respectively. These values are approximated by the dashed line at 50%. Error bars are 90% CI for the mean, also obtained from the empirical chance distribution.

Model and cortical generalization. Gaussian Bayes Naive, diagonal quadratic discriminant analysis (DQDA), linear support vector machine (SVM-lin), and non-linear SVM (SVM-RBF) classifiers were tested. The DQDA classifier was implemented with the Matlab fitcdiscr function, and SVMs were implemented with LIBSVM (Chang and Lin, 2011). SVM-lin used a linear kernel, and SVM-RBF used a radial-basis-function kernel with $\gamma = \frac{1}{\text{total ROI voxels}}$. The cost parameter was fixed at 10 for both SVM-lin and SVM-RBF.

The ROI data were submitted to the same LOSO cross-validation scheme used in the whole-brain analyses, and permutation testing was performed for statistical inference. For this permutation testing, class labels were randomly permuted, and ROI classification accuracies recalculated with the permuted class labels. This process was repeated 15 000 times, each with a consistent label permutation order which remained fixed across CV folds (matching the whole-brain analyses).

All four classifiers performed better than chance, and these results are shown in Figure 3. This confirms that the cortex identified by searchlight mapping is informative as a whole, not just as contiguous individual searchlights. Furthermore, this analysis demonstrates that our results generalize over multiple classification models.

Mean activity. Group differences in mean activation (e.g. regional univariate effects) could influence these results. To address this possibility, we repeated the previously described analyses after treating the data to control for mean activation. The fixed SVM-RBF cost was increased to 100 in this analysis, and all other model parameters were unchanged. First, voxel-wise 'cocktail blank' normalization (MacEvoy and Epstein, 2009) was performed on the *t*-value patterns within each run by setting each voxel to zero mean and feature scaling. After temporal compression, the same normalization was applied individually to each pattern across-voxels (i.e. spatially), so that all patterns had the same mean population activity. All four classifiers again performed better than chance on the treated data (Supplementary Figure S1), replicating the previous ROI follow-up analysis while controlling for mean activity. Furthermore, a whole-brain univariate analysis showed no group differences in mean bold activity for this cortical region (see Supplementary Material for univariate analysis details). These analyses confirm that the previous results were not simply driven by differences in mean activity; voxel patterns may be necessary, and are 'at least' sufficient for characterizing cortical specialization in this study [see Coutanche (2013) for greater detail on these analyses and their interpretations].

On a technical note, fixed SVM cost and γ parameters were chosen simply to avoid intensive parameter searches during cross-validation (Hsu *et al.*, 2003). Conducting such optimizations may have yielded superior performance, but the purpose of these analyses was to test generalization across different models, rather than comparing their relative performances.

Voxel alignment. Although the authors are not aware of any empirical findings, group differences in anatomical alignment could hypothetically influence between-groups voxel pattern analyses. In order to account for this concern analytically, another ROI analysis was performed with shared response modeling [SRM; Chen *et al.* (2015); also see Cohen *et al.* (2017) for a broader review].

This method maps subjects' fMRI data to a common latent variable space, based on subjects' shared responses to stimuli. SRM effectively performs inter-subject alignment in this way, aligning data in abstract response space rather than 3D anatomical space. This is similar to hyperalignment (Haxby *et al.*, 2011); however, SRM naturally facilitates between-groups analyses by estimating the latent representations shared by subjects[2] in order to find a common space. The ROI data were analyzed with the between-groups SRM analysis described in Chen *et al.* (2015) experiment 3, where group-specific responses are separated from both the common and idiosyncratic responses. These group-specific responses are then used for the between-groups LOSO classification. See Supplementary Material for further details.

This analysis successfully predicted participant's cultural identities (GNB accuracy = 97.5%; 90% CI for the mean = 82.94–100%, P < 0.001), and the SRM estimates substantially improved classification accuracy. This methodology should improve analytical sensitivity, as demonstrated in Chen *et al.* (2015). More importantly, this result shows culturally distinct visual cortex representations using methodology tolerant to anatomical misalignment. If American and East Asian voxel patterns conveyed the same stimulus information, but were anatomically misaligned, this analysis would have shown no distinction between

these groups. Instead, SRM estimations improved classification accuracy, providing good evidence that this study's results reflect differences in representational content.

Gender. Although there was no reason to expect gender differences in visual cortex representations, this study's participant groups were imbalanced with respect to gender. To address gender-imbalance concerns, two control analyses were conducted on the ROI data. In the first analysis, all male subjects were excluded, and the cross-validation scheme was altered to fairly account for those exclusions. This analysis successfully predicted participant's cultural identities (GNB accuracy = 88.99%; 90% CI for the mean = 80.15–98.97%, P < 0.001), while controlling for gender. The second analysis excluded male subjects from training, but included every male subject in the testing-sets (in addition to female subjects). This second analysis successfully predicted cultural identity as well (GNB accuracy = 75.74%; 90% CI for the mean = 68.82–83.14%, P < 0.001 p < .001). In both analyses the model was trained exclusively on female subjects, so the model's decision boundary could not discriminate cultural identity based on gender differences. The second analysis demonstrated that decision boundary also generalized to testing-sets comprised of both female and male participants. This indicates that this study's results are unlikely due to a gender-imbalance between cultural groups. The Supplemental Material describe these analyses in detail, including both ROC analyses and confusion matrices.

t-SNE. Improved methods for visualizing high-dimensional data provide valuable context for interpreting fMRI analyses (Johnson *et al.*, 2006; Peltonen and Kaski, 2011; Vellido Alcacena *et al.*, 2011). Therefore, the *t*-distributed stochastic neighbor embedding (t-SNE) algorithm was used to visualize the fMRI data submitted to the ROI follow-up analyses. t-SNE is a nonlinear dimension reduction technique which can produce 2D representations of high-dimensional data (Maaten and Hinton, 2008). Using this technique, a scatter plot was created illustrating how subjects' neural representations in visual cortex relate to each other (Figure 4a), and the quadratic discriminant analysis boundary was calculated for this 2D representation (Figure 4b). This figure is included purely for illustrative purposes, and is not quantitative evidence (see Supplementary Material for detailed methods).

Discussion

These results demonstrate culturally specialized visual cortex during a simple object viewing task. Additionally, this study is the first to show culturally distinct multi-voxel representations of stimuli that are non-specific to any particular culture. Raizada *et al.* (2010) found distinct fMRI pattern representations between Japanese and English participants, but in response to English syllables which have no functional distinction in Japanese. These stimuli are considerably different to Japanese and English speakers, so perhaps their dissimilar representation is less surprising. The stimuli in this experiment depicted objects familiar to both American and East Asian participants, yet the stimuli were nevertheless represented differently by American and East Asian participants in visual cortex.

This result reveals that one's cultural background specializes the cortex involved with object recognition, a routine and fundamental perceptual task. This finding is consistent with theories that predict prevalent and extensive cross-cultural

---

2 When compared with hyperalignment, which finds the rotations and reflections needed to make subjects' responses maximally similar.
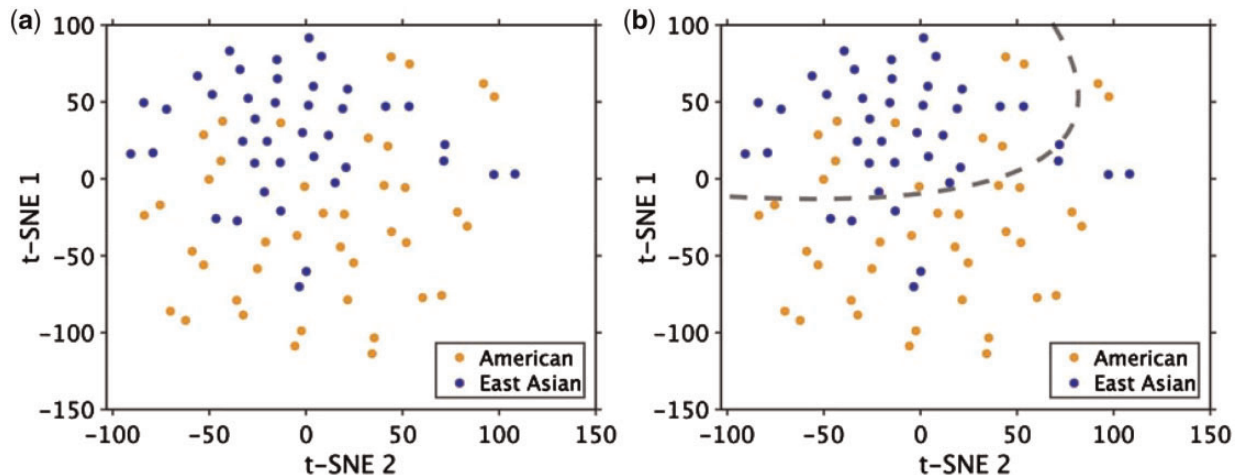
Fig. 4. t-SNE visualization of visual cortex representations. Each subject's ROI data is shown in 2D t-SNE space (a). The dashed line marks the QDA boundary (b). The visualized data consists of 80 total observations, with two ROI voxel patterns (one corresponding to each run) for each of the 40 subjects.

differences, such as differences in information processing between cultures. It is difficult to say whether these results reflect neuronal recycling, as the stimuli used in this study do not represent cultural inventions unique to either group. Neuronal recycling also proposes that humans' innate cortical organization limits how cultural acquisition may influence broader cognition. However, the real psychological limits imposed by these organizational constraints are unknown. In addition, this study cannot speak to whether culturally specialized visual cortex manifests as cross-cultural differences in information processing biases. Additional experiments would be needed to support that conclusion, as no experimental manipulations or analytical measures in this study addressed that specific hypothesis.

That being said, the current findings could coincide with recent work examining the cultural differences in information processing in light of spatial frequency tuning. Tardif *et al.* (2016) found Chinese participants utilized lower spatial frequency information than Canadian participants during a face recognition task [for further discussion of this idea, see Paige *et al.* (2017b)]. They concluded this difference in spatial frequency tuning underlies cultural differences in feature *vs* context information processing, because lower spatial frequency information is linked to narrower, feature-based focus (Shulman and Wilson, 1987). The cortex identified by searchlight mapping in the current experiment shows retinotopic sensitivity to spatial frequencies, that is, a strong cortical organization for responses to preferred spatial frequencies [see review by Tootell *et al.* (1998)]. Moreover, other research indicates that the region of visual cortex identified in this study responds to statistical regularities such as texture (Freeman *et al.*, 2013) and responds to the accumulation of high- and low-spatial frequency information (De Cesarei *et al.*, 2013). In addition, experimentally manipulating attention to high *vs* low spatial frequency consistently produces fMRI results with left-right hemispheric asymmetry (Han *et al.*, 2002; Iidaka *et al.*, 2004). It is plausible that cultural differences in spatial frequency tuning could explain why this study finds cultural specialization in left-lateralized visual cortex. Further research, employing stimuli designed to address this question, is needed (see Supplementary Material for a *post hoc* analysis).

A worthwhile future direction would be to directly test this hypothesis with classification image methods, where behavioral responses can be directly mapped to changes in stimulus

spatial frequency [see Murray (2011) for review]. Such an experiment would have ideal control over the stimuli's spatial frequency, while also providing a strong test of cultural differences by employing artificial stimuli that should not differ across cultures. Although this study's stimuli were rated equally familiar across cultures, it is impossible to ensure these stimuli were perfectly matched. However, Charest *et al.* (2014) found pictorial stimuli representations in early visual cortex (EVC; V1–V3) are both robustly common across people, and idiosyncratic to individuals. Furthermore, in their experiment, idiosyncratic EVC representations were surprisingly not observed when the stimuli were unequally familiar across participants (e.g. one participant's own house); that is, participant's EVC representations only differed when the stimuli were equally familiar to everyone. Our results indicate that such EVC representations are also systematically different between cultures.

Two other studies reporting culture-specific effects in this cortical area provide additional context for these results. Szwed *et al.* (2014) reported higher mean BOLD activity in left-lateralized BA 17–19 when Chinese and French participants read letter-scrambled words in their own native script. Gutchess *et al.* (2006) found dissimilar mean BOLD responses to focal objects *vs* contextual backgrounds between East Asian and American participants in left-lateralized BA 18 and 19 as well. This study involved a simple perceptual task, like this study, where participants viewed pictures of a central object within a contextual background (e.g. an elephant by a watering hole), and also pictures of those objects and backgrounds in isolation (e.g. only the elephant or watering hole). These studies show how the cultural differentiation of visual cortex can support either neuronal recycling or cultural differences in information processing. However, this study shows cultural specialization in visual cortex without an experimental manipulation specific to either theory. Given that these theories are not mutually exclusive, a more comprehensive hypothesis may unify these results.

One such hypothesis may be that culture-specific behavior specializes the cortex functionally relevant for that behavior, which in turn may broadly influence cognition beyond the scope of culturally acquired behavior. For example, reading in one's native script, an example of culture-specific behavior, would specialize visual cortex for reading that particular language. The specialization would tune this cortex for the spatial frequency information most useful for reading that language's

script (Horie *et al.*, 2012). Consequently, perceiving and recognizing objects, an example of non-culture-specific cognition, utilizes the visual cortex, which has been tuned for information on particular spatial frequencies. Language could potentially specialize visual cortex in a similar manner. Primary sensory cortices have shown multi-sensory coding (Liang *et al.*, 2013), and object representations have shown some consistency across modalities (Shinkareva *et al.*, 2011). The cortex performing object recognition may be culturally specialized through correlated auditory and visual stimuli, or higher-level associations with linguistic representations. This hypothesis would produce perceptual differences across cultures consistent with cultural differences in information processing, as may be the case in Tardif *et al.* (2016). Neuronal recycling may also accommodate this hypothesis, if the theoretical constraints proposed by Dehaene and Cohen (2007) are informed by evidence for broader perceptual differences. This hypothesis is far beyond this study's reach, although future studies along these lines may provide a more comprehensive explanation for perceptual differences across cultures.

Conclusion

This study successfully predicts participants' cultural identities from multi-voxel pattern representations of common objects in visual cortex. This result demonstrates the cultural specialization of visual cortex during a perceptual task ubiquitous to everyday life. Future research should address the fundamental mechanisms driving cultural cortex specialization, and how that specialization manifests downstream across behavioral domains. Attempting to answer these questions should further advance understanding of how culture shapes cognition.

Supplementary data

Supplementary data are available at *SCAN* online.

References

Baker, C.I., Liu, J., Wald, L.L., Kwong, K.K., Benner, T., Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(21), 9087–92.

Boduroglu, A., Shah, P., Nisbett, R.E. (2009). Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology*, **40**(3), 349–60.

Chang, C.-C., Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **2**(3), 1.

Charest, I., Kievit, R.A., Schmitz, T.W., Deca, D., Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(40), 14565–70.

Charest, I., Kriegeskorte, N. (2015). The brain of the beholder: honouring individual representational idiosyncrasies. *Language, Cognition and Neuroscience*, **30**(4), 367–79.

Chen, P.-H. (Cameron), Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., Ramadge, P.J. (2015). *A reduced-dimension fMRI shared response model*. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. editors. Advances in Neural Information Processing Systems 28. pp. 460–8 (Curran Associates, Inc., 2015).

Chiao, J.Y. (2009). Cultural neuroscience: a once and future discipline. *Progress in Brain Research*, **178**, 287–304.

Cohen, J.D., Daw, N., Engelhardt, B., *et al.* (2017). Computational approaches to fMRI analysis. *Nature Neuroscience*, **20**(3), 304–13.

Cohen, L., Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, **22**(1), 466–76.

Coutanche, M.N. (2013). Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cognitive, Affective, and Behavioral Neuroscience*, **13**(3), 667–73.

Dale, A.M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, **8**(2–3), 109–14.

De Cesarei, A., Mastria, S., Codispoti, M. (2013). Early spatial frequency processing of natural images: an ERP study. *PLoS One*, **8**(5), e65103.

Dehaene, S., Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, **56**(2), 384–98.

Eklund, A., Andersson, M., Josephson, C., Johannesson, M., Knutsson, H. (2012). Does parametric fMRI analysis with SPM yield valid results?—an empirical study of 1484 rest datasets. *NeuroImage*, **61**(3), 565–78.

Eklund, A., Nichols, T.E., Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(28), 7900–5.

Etzel, J.A., Zacks, J.M., Braver, T.S. (2013). Searchlight analysis: promise, pitfalls, and potential. *Neuroimage*, **78**, 261–9.

Evans, K., Rotello, C.M., Li, X., Rayner, K. (2009). Scene perception and memory revealed by eye movements and receiver-operating characteristic analyses: does a cultural difference truly exist? *Quarterly Journal of Experimental Psychology (Hove)*, **62**(2), 276–85.

Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., Movshon, J.A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, **16**(7), 974–81.

Gutchess, A.H., Indeck, A. (2009). Cultural influences on memory. *Progress in brain research*, **178**, 137–50.

Gutchess, A.H., Welsh, R.C., Boduroglu, A., PARK, D.C. (2006). Cultural differences in neural function associated with object processing. *Cognitive, Affective, and Behavioral Neuroscience*, **6**(2), 102–9.

Han, S., Ma, Y. (2014). Cultural differences in human brain activity: a quantitative meta-analysis. *NeuroImage*, **99**, 293–300.

Han, S., Weaver, J.A., Murray, S.O., Kang, X., Yund, E.W., Woods, D.L. (2002). Hemispheric asymmetry in global/local processing: effects of stimulus position and spatial frequency. *NeuroImage*, **17**(3), 1290–9.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, **293**(5539), 2425–30.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., *et al.* (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, **72**(2), 404–16.

Hedden, T., Ketay, S., Aron, A., Markus, H.R., Gabrieli, J.D. (2008). Cultural influences on neural substrates of attentional control. *Psychological Science*, **19**(1), 12–7.

Horie, S., Yamasaki, T., Okamoto, T., *et al.* (2012). Distinct role of spatial frequency in dissociative reading of ideograms and phonograms: an fMRI study. *Neuroimage*, **63**(2), 979–88.

Hsu, C.-W., Chang, C.-C., Lin, C.-J. (2003). A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Iidaka, T., Yamashita, K., Kashikura, K., Yonekura, Y. (2004). Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. An investigation with event-related fMRI. *Cognitive Brain Research*, **18**(2), 196–204.

Johnson, C., Moorhead, R., Munzner, T., Pfister, H., Rheingans, P., Yoo, T.S. (2006). NIH/NSF visualization research challenges report. Paper presented at the Los Alamitos, CA: IEEE Computing Society.

Kensinger, E.A., Garoff-Eaton, R.J., Schacter, D.L. (2007). Effects of emotion on memory specificity in young and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, **62**(4), P208–15.

Kitayama, S., Duffy, S., Kawamura, T., Larsen, J.T. (2003). Perceiving an object and its context in different cultures A cultural look at new look. *Psychological Science*, **14**(3), 201–6.

Kitayama, S., Uskul, A.K. (2011). Culture, mind, and the brain: current evidence and future directions. *Annu Rev Psychol*, **62**(1), 419–49.

Kriegeskorte, N., Goebel, R., Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(10), 3863–8.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, **12**(5), 535–40.

Liang, M., Mouraux, A., Hu, L., Iannetti, G.D. (2013). Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature Communications*, **4**(1), ncomms2979.

Maaten, L.V.D., Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, **9**, 2579–605.

MacEvoy, S.P., Epstein, R.A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, **19**(11), 943–7.

Madan, C.R. (2015). Creating 3D visualizations of MRI data: A brief guide [version 1; referees: 3 approved]. *F1000Research*, **4**, 466.

Masuda, T., Ellsworth, P.C., Mesquita, B., Leu, J., Tanida, S., Van de Veerdonk, E. (2008). Placing the face in context: cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, **94**(3), 365.

Masuda, T., Nisbett, R.E. (2001). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, **81**(5), 922–34.

Millar, P., Serbun, S., Vadalia, A., Gutchess, A. (2013). Cross-cultural differences in memory specificity. *Culture and Brain*, **1**(2–4), 138–57.

Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, **53**(1), 103–18.

Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., Brammer, M. (2006). The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *NeuroImage*, **33**(4), 1055–65.

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, **59**(3), 2636–43.

Murray, R.F. (2011). Classification images: a review. *Journal of Vision*, **11**(5), 2.

Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, **108**(2), 291.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*, **10**(9), 424–30.

Paige, L.E., Amado, S., Gutchess, A.H. (2017a). Influence of encoding instructions and response bias on cross-cultural differences in specific recognition. *Culture and Brain*, **5**(2), 153–68.

Paige, L.E., Ksander, J.C., Johndro, H.A., Gutchess, A.H. (2017b). Cross-cultural differences in the neural correlates of specific and general recognition. *Cortex*, **91**, 250–61.

Paulesu, E., McCrory, E., Fazio, F., *et al.* (2000). A cultural effect on brain function. *Nat Neurosci*, **3**(1), 91–6.

Peltonen, J., Kaski, S. (2011). Generative modeling for maximizing precision and recall in information visualization. Paper presented at the AISTATS.

Pereira, F., Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. *Neuroimage*, **56**(2), 476–96.

Raizada, R.D., Lee, Y.-S. (2013). Smoothness without smoothing: why Gaussian naive Bayes is not naive for multi-subject searchlight studies. *PLoS One*, **8**(7), e69566.

Raizada, R.D., Tsao, F.-M., Liu, H.-M., Kuhl, P.K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences. *Cerebral Cortex*, **20**(1), 1–12.

Rule, N.O., Freeman, J.B., Ambady, N. (2013). Culture in social neuroscience: a review. *Society for Neuroscience*, **8**(1), 3–10.

Serences, J.T., Boynton, G.M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, **55**(2), 301–12.

Shinkareva, S.V., Malave, V.L., Mason, R.A., Mitchell, T.M., Just, M.A. (2011). Commonality of neural representations of words and pictures. *Neuroimage*, **54**(3), 2418–25.

Shulman, G.L., Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, **16**(1), 89–101.

Stelzer, J., Chen, Y., Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage*, **65**, 69–82.

Szwed, M., Qiao, E., Jobert, A., Dehaene, S., Cohen, L. (2014). Effects of literacy in early visual and occipitotemporal areas of

Chinese and French readers. *Journal of Cognitive Neuroscience*, **26**(3), 459–75.

Tardif, J., Fiset, D., Zhang, Y., *et al.* (2016). Culture shapes spatial frequency tuning for face identification. *Journal of Experimental Psychology. Human Perception and Performance*, **43**(2), 294.

Tootell, R.B., Hadjikhani, N.K., Mendola, J.D., Marrett, S., Dale, A.M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences*, **2**(5), 174–83.

Vellido Alcacena, A., Martín, J.D., Rossi, F., Lisboa, P.J. (2011). Seeing is believing: The importance of visualization in real-world machine learning applications. Paper presented at the Proceedings: 19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2011: Bruges, Belgium, April 27–29, 2011.

Wang, H., Masuda, T., Ito, K., Rashid, M. (2012). How much information? East Asian and North American cultural products and information search performance. *Personality and Social Psychology Bulletin*, **38**(12), 1539–51.

Zhou, J., Gotch, C., Zhou, Y., Liu, Z. (2008). Perceiving an object in its context: is the context cultural or perceptual? *Journal of Vision*, **8**(12), 2–5.