**Supplemental Methods**

**Temporal compression**

For the final preprocessing step, run-wise temporal compression was performed on all valid searchlight voxels. In this procedure, the trial HDR estimates were averaged within each run, producing two voxel response patterns per subject (one pattern per run). Temporal compression improves signal-to-noise ratio without spatial smoothing by averaging over noisy trial-to-trial pattern variance (Mourao-Miranda, Reynaud, McGlone, Calvert, & Brammer, 2006). In the current study, this transformation also aggregates representations across different visual stimuli in order to detect a perceptual neural code that is distinct between culture groups, but also common across stimuli. If culture specializes the cortex responsible for perception, then different stimuli should reliably produce representations which reflect specialization. Aggregating information across trials should improve the signal for such specialization, while also suppressing the irrelevant trial-to-trial variance from information related to the particulars of different visual stimuli. These temporally compressed t-value patterns were the input data for searchlight classification.

**Voxel exclusion**

We set these exclusion criteria a priori to minimize spatial inconsistency in our whole brain analysis, while also minimizing the impact of missing data. Trials missing certain voxel values for a particular subject would create different testing and training sets across the brain, given our LOSO cross-validation scheme. Such a scenario where cross-validation sets differ (potentially dramatically) with an unknown spatial distribution across a large number of voxels would be highly difficult to account for analytically. In total, a small proportion of voxels were excluded. 198,071 voxels survived gray matter thresholding and 3,274 voxels were subsequently excluded for dropout (1.65% rejection rate). All dropout-exclusion voxels were located in extreme inferior cerebellum and brainstem, outside or near the image acquisition bounding-box in our scanning protocol.

**Searchlight mapping statistical inference**

The non-parametric cluster significance test developed by Stelzer, Chen, and Turner (2013) was adapted for the current between-subjects MVPA design. Class labels were randomly permuted, and searchlight accuracy maps were recalculated with the permuted class labels. Class relabeling was performed without constraints regarding subject-grouping, so that subjects were allowed to have heterogeneously relabeled datapoints. This process was repeated 15,000 times, producing 15,000 whole brain searchlight maps. Each of these searchlight maps used a consistent label permutation order, and this order also remained fixed across CV folds (see Stelzer et al. (2013) for additional detail on these design parameters).
First, accuracy thresholds were calculated for every voxel\(^1\). The combined 15,000 permuted searchlight maps produced an empirical chance-accuracy distribution at every voxel location. Each voxel’s accuracy threshold was calculated as the accuracy needed to reject a chance null hypothesis at \(\alpha = .001\) against the voxel’s empirical chance distribution.

Next, a null cluster size distribution was created. This was accomplished by finding the above-threshold accuracy searchlight clusters that occurred during permutation testing. Within every permuted searchlight map, first the voxels (i.e. searchlight centers) with above-threshold accuracies were identified. A cluster search algorithm then located clusters of contiguous above-threshold voxels. This search used a six-connectivity scheme, which considered voxels contiguous if they share a face. Finally, the cluster sizes found in each permuted searchlight map were aggregated into one distribution.

The same voxel-wise thresholding and cluster search procedures were applied to the original searchlight map obtained from the real class labeling in this experiment. The \(p\)-values for each cluster found in the real data were calculated as \(P_{\text{cluster}} = p(S > H)\), where \(S\) is the cluster’s size and \(H\) is the null cluster size distribution. A step-down false discovery rate (FDR) correction was then applied to these cluster \(p\)-values (Benjamini & Liu, 1999).

This adaptation only deviated from the original Stelzer et al. (2013) procedure in one major regard, in that 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 the current study were analyzed in the same fashion as the bootstrapped group accuracy maps in Stelzer et al. (2013).

**Whole-brain univariate analysis**

In order to determine whether our whole-brain searchlight results also reflected group differences in mean BOLD activity, we ran an analogously designed whole-brain univariate analysis. Using the SPM12 toolbox, the BOLD scans were first smoothed using a 6-mm FWHM Gaussian kernel. Subjects’ hemodynamic responses to all stimuli were modeled with a single regressor at the first-level; these responses were then analyzed at the group-level by treating subjects as a random effect. Group differences were assessed with t-tests thresholded at \(p<.001\) and \(k=10\) voxel extent.

With these liberal thresholding parameters, we observed no supra-threshold clusters in the vicinity of our searchlight results. Our ROI testing indicated that voxel patterns are at least sufficient for characterizing cortical specialization in this study, and these univariate results suggest they may be necessary as well (see Coutanche (2013) for greater detail on these analyses and their interpretations).

**Voxel alignment**

Group differences in anatomical alignment could hypothetically influence between-groups voxel pattern analyses. Although the authors are not aware of any empirical findings on

\(^1\) Classification accuracies at each voxel position correspond to the searchlight centered at that position.
this topic, the current study could have plausibly been influenced in this way. In order to account for this concern analytically, the ROI data was analyzed 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. We employed 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.

First, the shared response common to all subjects \(S^{all}\) is found with data from each subject \(i\) so that:

\[
X_i = W_i S^{all} + E_i^{all}
\]

(1)

where \(X\) is the \((v \text{ voxels} \times d \text{ stimuli})\) fMRI data matrix, \(W\) is the \((v \times k\) latent features) subject-specific base matrix, \(S\) is the \((k \times d)\) shared response matrix, and \(E\) is the \((v \times d)\) residual matrix. The \(E_i^{all}\) term provides the subject responses that do not include the latent response shared by all subjects \(S^{all}\).

Next, the residuals from equation 1 are used to estimate the group-specific response \(S^{gj}\) for all subjects \(i\) in group \(j\):

\[
E_i^{all} = W_i^{gj} S^{gj} + E_i^{gj}
\]

(2)

This gives group-specific responses now separated from the common response \(S^{all}\), and the idiosyncratic responses \(E_i^{gj}\) not captured by either \(S^{all}\) or \(S^{gj}\). Finally, these group-specific responses are mapped back into subjects’ voxel space:

\[
X_i^{gj} = W_i^{gj} S^{gj}
\]

(3)

We obtained the voxel space data \(X_i^{gj}\) for \(k = 10\) latent features from our visual cortex ROI data. This SRM voxel space data was then tested with the same follow-up analysis protocol referenced in Figure 3 of the manuscript. This analysis successfully predicted participant’s cultural identities (GNB accuracy = 97.5%; 90% C.I. for the mean = 82.94% - 100%, \(p < .001\)), and the SRM estimates substantially improved classification accuracy.

**Gender**

Although there was no reason to expect gender differences in visual cortex representations, the current study’s participant groups were not well balanced with respect to gender. To address gender-imbalance concerns, control analyses were run on the ROI identified by searchlight mapping. In the first analysis, all male subjects were excluded and the cross-validation scheme was altered to fairly account for those exclusions. In this scheme, 1000 cross-validation folds were created by randomly selecting a subset of subjects for the fold’s training-

\(^2\) Compared to hyperalignment, which finds the rotations and reflections needed to make subjects’ responses maximally similar.
set (no two folds contained the same set). The ratio of East Asian to American subjects within each training-set matched the original LOSO analysis’ ratio. Specifically, one group always had 10 subjects, and the other group had 9 subjects (the LOSO ratio was 20:19 in the main analyses). This was the maximum number of subjects which could be selected for training, since there were only 10 female East Asian participants. Within each fold, the remaining female subjects comprised the testing-set (n = 8). The classification accuracy across cross-validation folds was tested for significance with the same permutation testing method described in the ROI follow-up analyses section.

This analysis successfully predicted participant’s cultural identities (GNB accuracy = 88.99%; 90% C.I. for the mean = 80.15% - 98.97%, p < .001), while controlling for gender. A second analysis followed the same procedure, but each testing-set also included all male subjects (increasing the testing-set to n = 21). This second analysis successfully predicted cultural identity as well (GNB accuracy = 75.74%; 90% C.I. for the mean = 68.82% - 83.14%, 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.

ROC curves were also evaluated for both gender control analyses, as shown in supplementary figure 5. Hypothesis testing was conducted with area under the curve (AUC) statistics, showing greater than chance performance with this metric as well. This was true for both the control analyses with exclusively-female testing sets (AUC = .73; 90% C.I. for the mean = .61 - .86, mean empirical chance AUC = .47, p = .003) and mixed-gender testing sets (AUC = .86; 90% C.I. for the mean = .77 - .96, mean empirical chance AUC = .49, p < .001). Supplementary figure 6 shows also the confusion matrices for these analyses.

Of note, the improvement in ROC curve behavior by including male subjects in the testing-set can be explained by the testing-set’s larger sample size (n=21, instead of n=8) and more balanced class ratio (i.e. ratio of American to East Asian datapoints). The same is true for the confusion matrices in Supplementary Figure 6, as those statistics similarly describe analytical sensitivity and power. In panel A the model predicts “US” much more often (top row, green and red) because the exclusively female testing-set contains far more Americans than East Asians. When males are included in the testing-set (panel B), the testing-set contains a more balanced ratio of Americans and East Asians, and the model predictions change accordingly.

**t-SNE visualization**

A t-SNE perplexity parameter search was performed for this visualization, similar to the approach described by Mwangi, Soares, and Hasan (2014). The optimal perplexity parameter (32) was determined by the highest global silhouette index obtained via k-means clustering with k = 2, as this analysis concerns two subject groups. The two dimensional data representation produced by the optimal perplexity parameter was then plotted (manuscript
Spatial frequency RSA

In the discussion, we speculated that our results may have been produced by cultural differences in spatial frequency. Unfortunately, the current study was not designed for directly testing hypotheses regarding particular stimulus properties. With the caveat that strong conclusions cannot be drawn from inappropriate tests (as we did not a priori manipulate potentially relevant visual properties), we conducted another follow-up analysis examining group differences in spatial frequency tuning, as this was one potential mechanism that we identify as of interest. We performed a representational similarity analysis (RSA; see Kriegeskorte, Mur, and Bandettini (2008) for an overview) on the visual cortex described in ROI follow-up analyses. Our RSA closely mirrored the analysis of spatial frequency coding in early visual cortex conducted by Kriegeskorte et al. (2008). This analysis compares the stimuli correlation structure to the correlation structure of voxel patterns evoked by those stimuli. A strong coherence between these correlation structures indicates the fMRI data reflects stimuli information (i.e. codes for the stimuli). In our analysis, we also compared this measure of coherence between Americans and East Asians.

The original color stimuli were first converted to grayscale images (supplementary figure 2 panel a), then we isolated the low and high spatial frequency characteristics of our stimuli by producing low-pass and high-pass filtered versions of these images. Low-pass stimuli (supplementary figure 2 panel b) were produced by convolving the 256 x 256 pixel grayscale images with a Gaussian kernel (σ = 7.3), and then high-pass stimuli (supplementary figure 2 panel c) were calculated by subtracting those low-pass images from their original grayscale counterparts.

We computed two representational dissimilarity matrices (RDMs) characterizing the low and high spatial frequency information in our stimuli (supplementary figure 3, panels A and B respectively). The low spatial frequency information RDM was constructed by converting each low-pass image to a pixel vector, and then calculating the pair-wise rank-correlation distance (1 – *Spearman’s rho*) between all pixel vectors. The high spatial frequency information RDM was constructed from the high-pass images in exactly the same way.

Visual cortex RDMs were computed for each subject by vectorizing the ROI voxel HDRs estimated for each stimulus (HDR estimation described in Hemodynamic Response Modeling), and then calculating the pair-wise rank-correlation distance between all stimuli HDR vectors. Accordingly, trial HDR estimates were not temporally compressed in this analysis. The ordering of RDM entries was consistent across all visual cortex and spatial frequency information RDMs, so that corresponding matrix entries referred to the same stimulus. We formed two group RDMs (American and East Asian) by simply averaging subjects’ visual cortex RDMs within each group.
This operation provides two advantages: evidence is combined later in the analysis, and subjects’ voxel patterns are never directly compared (see Voxel Alignment in the manuscript).

In our RSA, we first calculated visual cortex RDMs within each subject by correlating their voxel response patterns to each stimulus; this results in matrices that reflect the correlation structure of their stimulus representations. We then averaged and compared these RDMs at a second level removed from subjects’ voxel response patterns (i.e. correlations of correlation matrices, rather than correlations of voxel patterns).

We compared these group visual cortex RDMs with the spatial frequency information RDMs in order to examine how well Americans and East Asians coded for stimuli spatial frequency (within our visual cortex ROI). In the first comparisons, we determined whether visual cortex RDMs significantly coded for high and low frequencies by themselves. We tested this by correlating group visual cortex RDMs with spatial frequency RDMs, then testing the significance of those correlations via permutation testing (permuting the entries of one RDM 15,000 times, within each comparison). As shown in supplementary figure 4, Americans coded for both high and low spatial frequency (high: \( r = .04, p = .03 \); low: \( r = .07, p < .01 \)), whereas East Asians did not significantly code for either (high: \( r = .02, p = .32 \); low: \( r = -.003, p = .37 \)). We then formally tested whether East Asians and Americans coded for spatial frequency differently. For each spatial frequency RDM (high, then low), we tested the difference between East Asian and American visual cortex RDM fits via the stimuli-set bootstrapping method described in Kriegeskorte et al. (2008). In this procedure, the stimuli set was resampled 15,000 times. We found East Asians and Americans coded differently for low spatial frequency (\( p = .048 \)), but not high spatial frequency (\( p = .28 \)); these results are also shown in supplementary figure 4. As a side note, we also performed a variation of this analysis where we did not create group average RDMs, but instead tested individual subjects’ RDMs and evaluated the group mean fits. We obtained the same pattern of results from that analysis as well.

These results provide some evidence supporting our speculation that cultural differences in spatial frequency tuning produced our main result. The RSA indicates that Americans and East Asians coded spatial frequency differently during our experiment. Although we might have predicted that Americans were more attuned to high spatial frequencies and East Asians to low spatial frequencies, the results do not clearly support that interpretation. Cultural differences were more pronounced for low-pass filtered stimuli, for which Americans and East Asians showed significantly different coding. At this point, it is difficult to link a greater tendency to code for low vs. high spatial frequencies to cultural differences in behavioral outcomes, and whether more robust coding reflects greater reliance on the information, or more ease or difficulty of processing, for example. Further work is needed to aid in the interpretation of this finding, as well as substantiate the pattern in an experiment designed a priori for such an analysis. The importance of spatial frequency coding in this context would also be substantiated by a behavioral task, but our present experiment was not designed with that in mind.
Supplementary Figure 1. ROI analyses controlling for mean activation. The mean empirical chance accuracies were 48.71%, 48.8%, 49.19%, and 49.18% for GNB, DQDA, SVM-lin, and SVM-RBF, respectively. These values are approximated by the dashed line at 50%. Error bars are 90% C.I. for the mean, also obtained from the empirical chance distribution.
Supplementary Figure 2. Stimuli examples. Grayscale image (A), low-pass (B), and high-pass (C).

Supplementary Figure 3. Stimuli spatial frequency RDMs.
**Supplementary Figure 4.** RSA results. Significant spatial frequency coding is noted by (*); significant coding differences between groups are noted by (†). Error bars show standard errors for the deviation estimates (standard deviation of RDM fits from bootstrap resampling the stimuli-set).

**Supplementary Figure 5.** ROC curves for gender control analyses. Exclusively female testing sets are shown in panel A (AUC = .73; 90% C.I. for the mean = .61 - .86, mean empirical chance AUC = .47), and mixed-gender testing sets are shown in panel B (AUC = .86; 90% C.I. for the mean = .77 - .96, mean empirical chance AUC = .49).
Supplementary Figure 6. Confusion matrices for gender control analyses. Exclusively female testing sets are shown in panel A, and mixed-gender testing sets are shown in panel B.
Supplementary Table 1.
*Searchlight center voxel locations.*

| Brodmann Area | MNI coordinates |   |   |   |
|---------------|----------------|---|---|---|
| 18            | -28            | -94| 18|
| 18            | -26            | -94| 18|
| 18            | -28            | -92| 18|
| 18            | -24            | -96| 20|
| 18            | -26            | -94| 20|
| 18            | -24            | -94| 20|
| 18            | -26            | -92| 20|
| 18            | -24            | -92| 20|
| 18            | -24            | -94| 22|
| 18            | -24            | -92| 22|
| 18            | -24            | -92| 24|
| 18            | -22            | -92| 24|
| 18            | -24            | -90| 24|
| 18            | -22            | -90| 24|
| 18            | -22            | -92| 26|
| 18            | -20            | -92| 26|
| 18            | -18            | -92| 26|
| 18            | -30            | -90| 26|
| 18            | -28            | -90| 26|
| 18            | -26            | -90| 26|
| 18            | -24            | -90| 26|
| 18            | -22            | -90| 26|
| 18            | -20            | -92| 28|
| 18            | -18            | -92| 28|
| 18            | -28            | -90| 28|
| 18            | -22            | -90| 28|
| 18            | -18            | -92| 30|
| 18            | -16            | -92| 30|
| 18            | -18            | -90| 30|
| 19            | -32            | -90| 18|
| 19            | -36            | -88| 18|
| 19            | -34            | -88| 18|
| 19            | -34            | -90| 20|
| 19            | -32            | -90| 20|
| 19            | -34            | -88| 20|
| 19            | -32            | -88| 20|
| 19            | -32            | -90| 22|
| 19            | -32            | -90| 24|
| 19            | -30            | -90| 24|
| 19            | -32            | -88| 24|
| 19            | -28            | -88| 26|
References

Benjamini, Y., & Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of Statistical Planning and Inference, 82*(1), 163-170.

Chen, P. C., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. (2015). A reduced-dimension fMRI shared response model. Paper presented at the Advances in Neural Information Processing Systems.

Cohen, J. D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., . . . Turk-Browne, N. B. (2017). Computational approaches to fMRI analysis. *Nature Neuroscience, 20*(3), 304-313.

Coutanche, M. N. (2013). Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cognitive, Affective, & Behavioral Neuroscience, 13*(3), 667-673.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., . . . Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron, 72*(2), 404-416.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2.*

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), 1055-1065.

Mwangi, B., Soares, J. C., & Hasan, K. M. (2014). Visualization and unsupervised predictive clustering of high-dimensional multimodal neuroimaging data. *Journal of neuroscience methods, 236*, 19-25.

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.