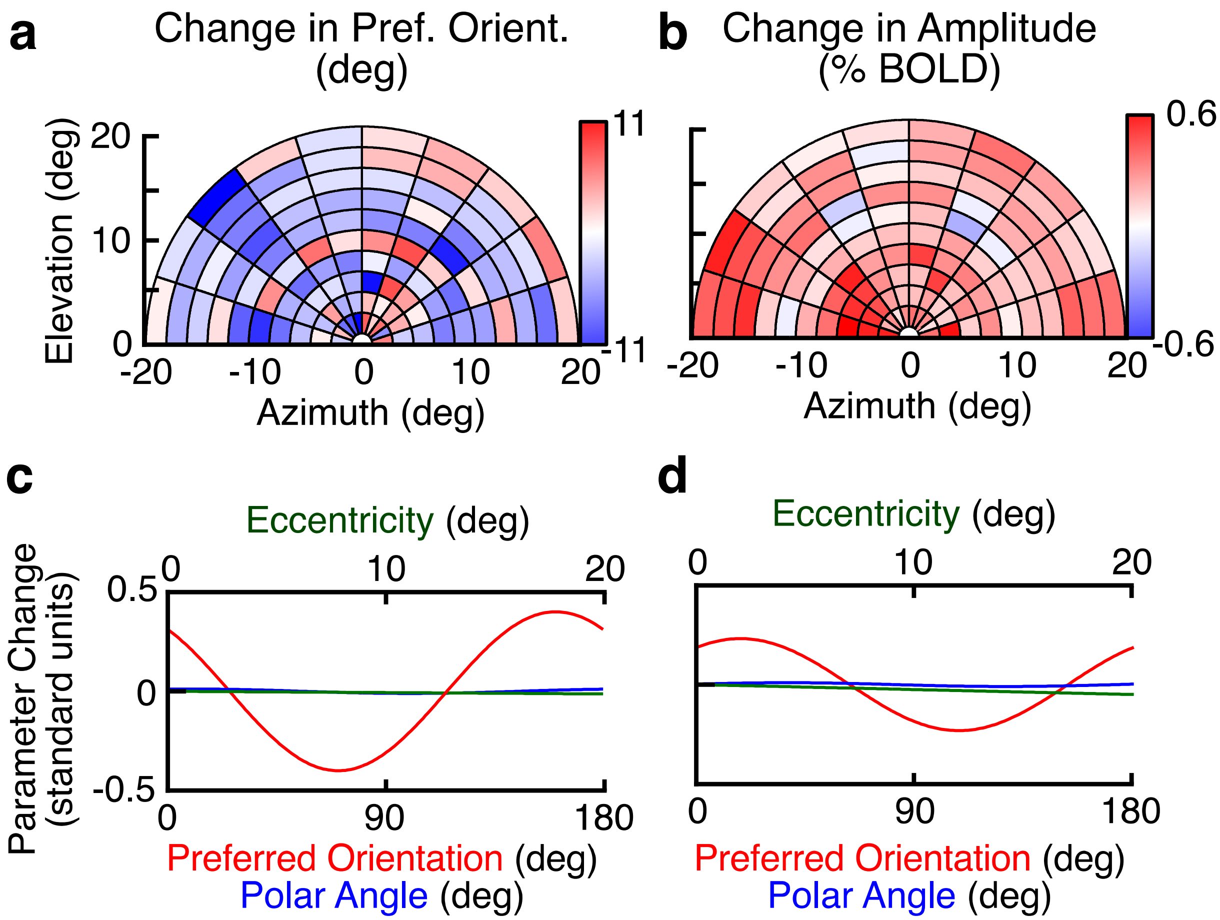
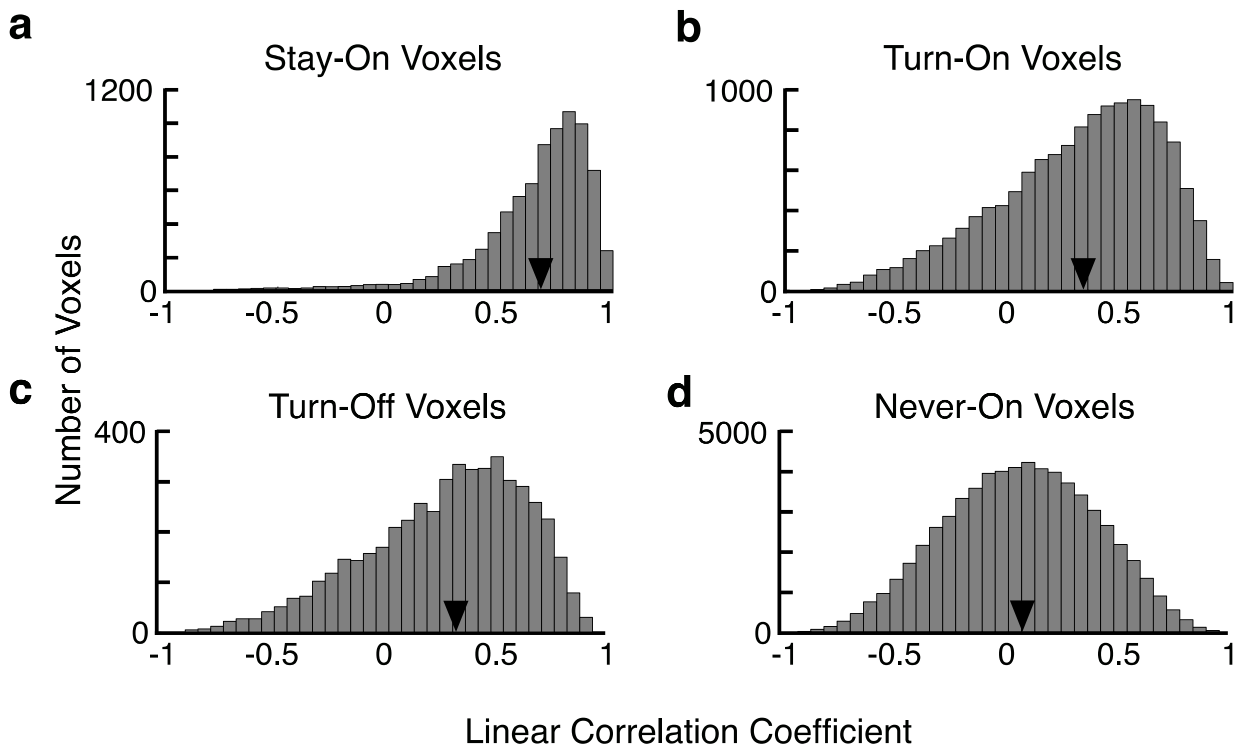
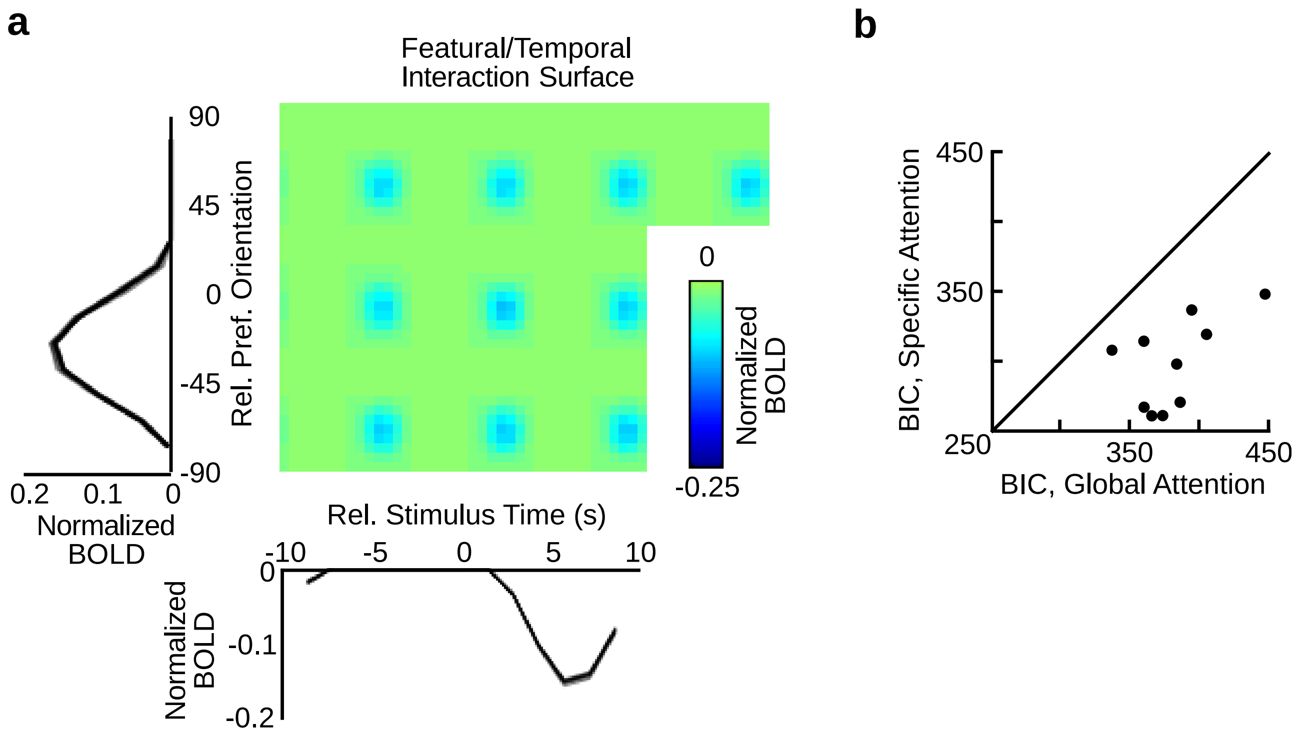


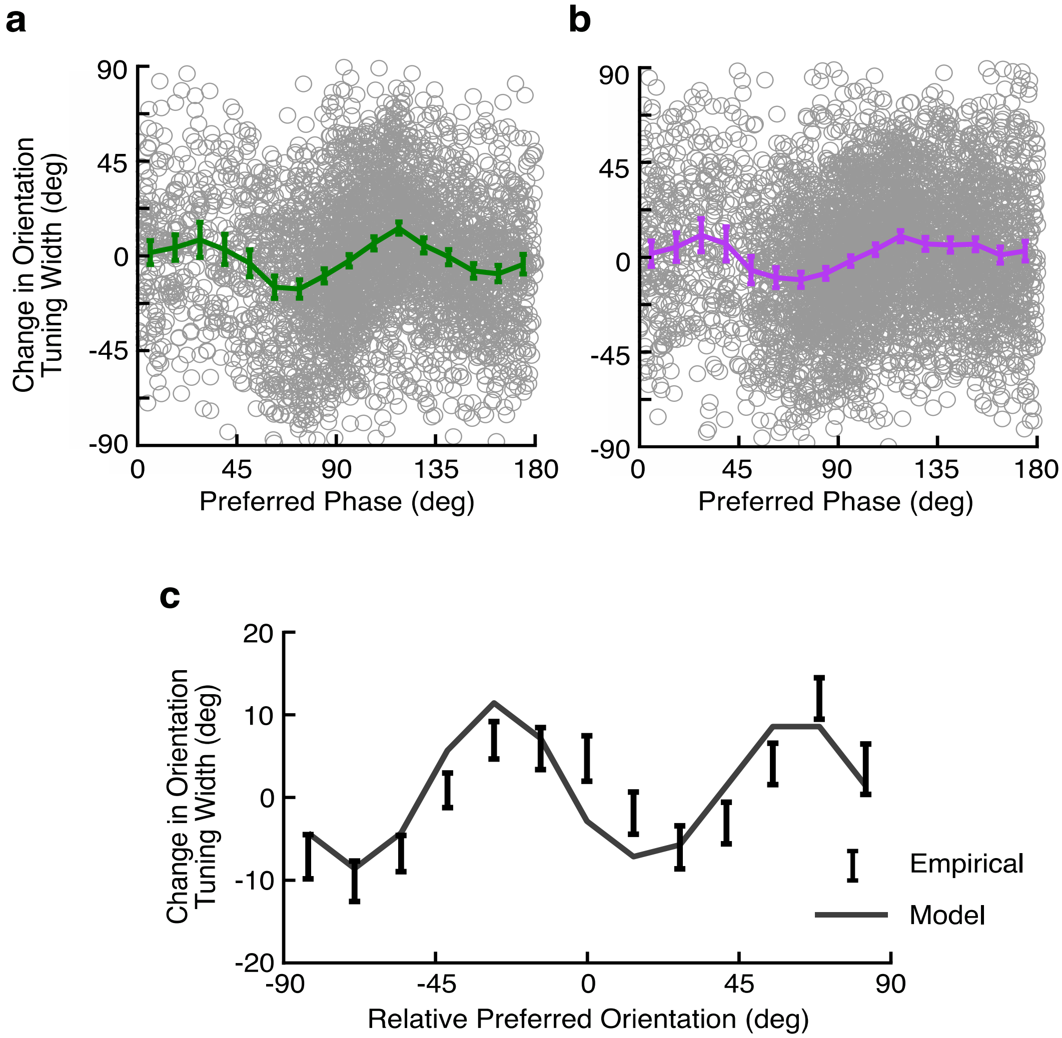
Supplementary Figure 1. Attentional effects are invariant over visual space. **A.** Changes in preferred orientation between the A45 and A135 attention conditions (positive number corresponds to larger values during A45 condition) are distributed randomly across visual space. **B.** Aside from a bias toward larger responses in A45, changes in amplitude are similarly random. **C, D.** Standardized regression attempts to predict changes in preferred orientation (C) and amplitude (D) simultaneously as functions of mean preferred orientation and retinotopic map location. The relative amplitude of the fitted functions indicates the relative amount of variance for which each predictor accounts. Neither polar angle nor eccentricity is able to explain the variations in preferred orientation or amplitude seen between the different attention conditions.



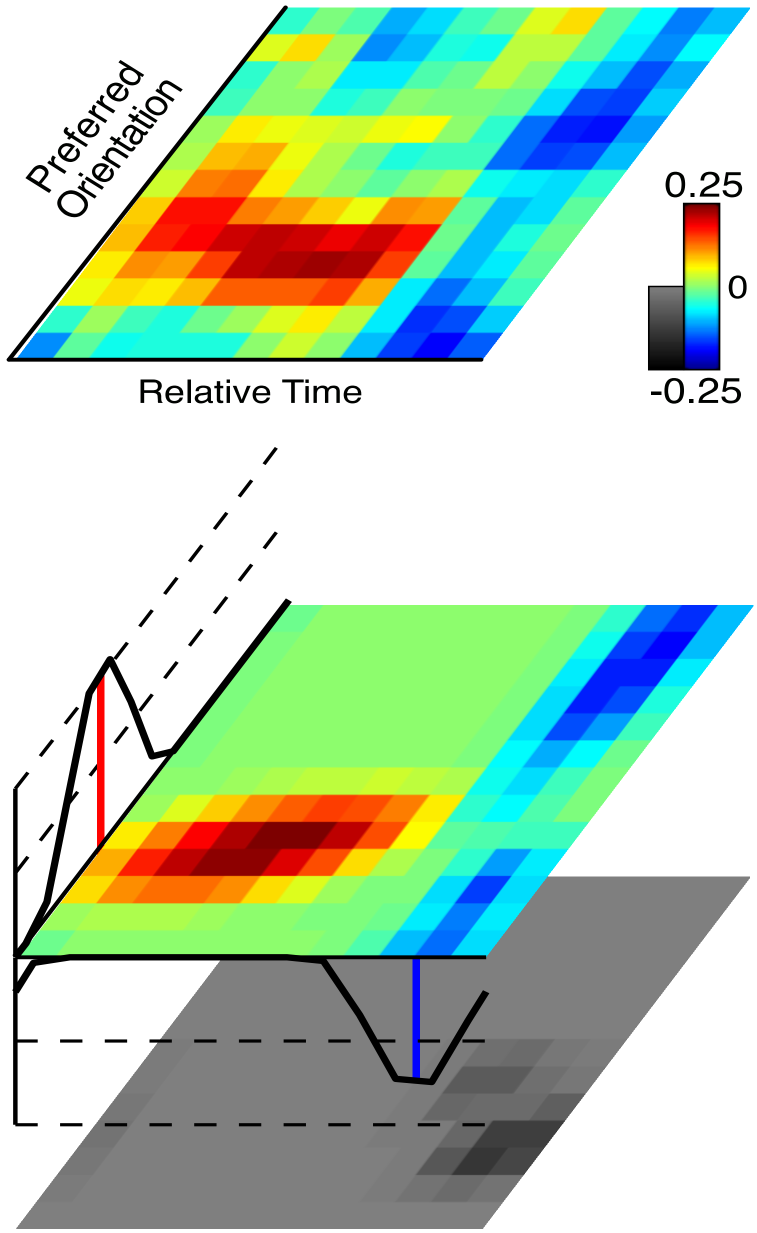
Supplementary Figure 2. Individual voxels show a linear relationship across task conditions. **A.** The Pearson's correlation coefficient between the tuning curves derived from the No-Cue and two Attention conditions is high for Stay-On voxels that were active during both conditions. *arrowhead*, median correlation coefficient is 0.71. **B, C.** For Turn-On and Turn-Off voxels, a lower but still positive coefficient is observed, showing that there is still a linear relationship between the tuning curves obtained when a voxel is and is not significantly periodic. (Turn-On median 0.35, Turn-Off median 0.33) **D.** Never-On voxels were not periodic in any task condition. These voxels provide a baseline measurement of the linear correlations expected by chance in our dataset (median 0.07).



Supplementary Figure 3. 90/10 Cross-validation of the Stay-On voxel attention model. Both a global-term Null model (only containing normalization terms) and the 9-parameter Attention model identified in the main paper were fitted to attention surfaces generated from 90% of all voxels, and the likelihood (represented as a Bayesian information criterion, BIC) of these models explaining the surfaces generated from the remaining 10% of voxels was computed. 10 repetitions total were performed: each voxel was used as training data 9 times and testing data once. **A.** The featural, temporal, and feature-time interaction terms obtained from all 10 repetitions are in excellent agreement. The 10 curves on each axis are indistinguishable. **B.** For all 10 repetitions, the full model with specific attention terms (y-axis) provides a superior explanation (lower BIC) of the testing data than does the non-specific null model, even when the seven-parameter difference between these models is taken into account.



Supplementary Figure 4. Attention changes the width of orientation tuning curves. Tuning width was estimated by fitting a circular Gaussian function that was modified to fit widths wider than 90˚ (see Methods). **A.** Compared to width measured in the No-Cue condition, widths during A45 follow a complex bimodal distribution. *line,* mean change in width versus preferred phase, *error bars* 95% confidence interval for the mean. **B.** During A135, a similar distribution of tuning width changes is observed. Note that, although there is substantial noise in the width changes, the bimodal curves are similar in shape with respect to the focus of attention (A: A45 and B: A135). **C.** Mean changes from A and B are aligned (error bars), and compared to predicted changes in tuning width derived from the attention model for Stay-On voxels (solid line, model shown in Figure 6). The model recovers the overall bimodal distribution of width changes.



Supplementary Figure 5. Gain-based models provide a satisfactory, but inferior, explanation of attentional effects. The attention surface (top) and the best-fitting multiplicative model (bottom) are shown in the same manner as in Figure 7. The three-term gain model (Featural gain, Temporal gain, Feature-Time Interaction gain) is the lowest-BIC model and all parameters are justified, just as seen in the additive model family. However, the gain models are inferior: BIC_{Additive} = 165 vs. BIC_{Gain} = 256, *weighted* SSE_{Additive} = 2.26 vs. $wSSE_{\text{Gain}}$ = 3.99, $R^2_{\text{adj Additive}}$ = 0.82 vs. $R^2_{\text{adj Gain}}$ = 0.67.