Attention and biased competition in multi-voxel object representations

Leila Reddy[a,b], Nancy G. Kanwisher[c], and Rufin VanRullen[a,b,1]

[a]Université de Toulouse, Université Paul Sabatier, Centre de Recherche Cerveau et Cognition, 31062 Toulouse, France; [b]Centre National de la Recherche Scientifique, CerCo, 31062 Toulouse, France; and [c]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02142

Edited by Robert Desimone, Massachusetts Institute of Technology, Cambridge, MA, and approved October 27, 2009 (received for review July 02, 2009)

The biased-competition theory accounts for attentional effects at the single-neuron level: It predicts that the neuronal response to simultaneously-presented stimuli is a weighted average of the response to isolated stimuli, and that attention biases the weights in favor of the attended stimulus. Perception, however, relies not on single neurons but on larger neuronal populations. The responses of such populations are in part reflected in large-scale multivoxel fMRI activation patterns. Because the pooling of neuronal responses into blood-oxygen-level-dependent signals is nonlinear, fMRI effects of attention need not mirror those observed at the neuronal level. Thus, to bridge the gap between neuronal responses and human perception, it is fundamental to understand attentional influences in large-scale multivariate representations of simultaneously-presented objects. Here, we ask how responses to simultaneous stimuli are combined in multivoxel fMRI patterns, and how attention affects the paired response. Objects from four categories were presented singly, or in pairs such that each category was attended, unattended, or attention was divided between the two. In a multidimensional voxel space, the response to simultaneously-presented categories was well described as a weighted average. The weights were biased toward the preferred category in category-selective regions. Consistent with single-unit reports, attention shifted the weights by ≈30% in favor of the attended stimulus. These findings extend the biased-competition framework to the realm of large-scale multivoxel brain activations.

pattern classification | fMRI | response combination | ventral temporal cortex

There has recently been an explosion of fMRI studies using multivariate patterns of blood-oxygen-level-dependent (BOLD) signals, distributed over large numbers of voxels, to probe neural representations and their relation to perception. This exciting body of work has shown that there is a wealth of information about the perceptual and cognitive states of the observer to be gained from such large-scale multivariate representations that would have been otherwise hidden in the univariate (i.e., average) BOLD response. For instance, decoding or "mind-reading" studies have shown that the visual stimulus (1–7), the contents of perceptual experience (8), working memory (9), or mental imagery (10) can all be predicted from the multivariate BOLD response. That is, large-scale multivariate fMRI patterns offer a previously unavailable window into human perceptual processes (11, 12).

Although attention has long been explored with conventional (i.e., univariate) fMRI methods (13–15), these recently discovered multivariate pattern analysis techniques have only started to scratch the surface of attentional processes. For example, it has recently been shown that one can decode which of two orientations (4) or motion directions (16) is currently attended, based on multivariate patterns in early striate and extrastriate cortex. However, the operational mechanisms of attention, which have been thoroughly explored in animal single-unit electrophysiology, leading to the well-established "biased-competition" framework (17), have yet to be understood at the level of multivariate fMRI patterns. Because of the robust correspondence between these large-scale patterns and human perceptual and cognitive variables, this understanding constitutes a necessary and critical step in bridging the gap between cognitive studies of attention and the detailed implementation of attentional mechanisms at the level of neuronal populations.

Extensive single-neuron recording studies in monkeys have revealed the effects of attention and competition on neural responses (18–23). When two stimuli, one effective and the other ineffective, are presented within a receptive field (RF) of a neuron, the neural response corresponds to a weighted average of the responses to the two stimuli presented in isolation (21). Attending to one or the other stimulus biases the neural response in favor of the attended stimulus. In the extreme case, the result is almost as if attention eliminates the influence of the unattended stimulus (18); in practice the bias, when quantified, is only ≈30% on average (21, 23–27). These results form the basis of the biased-competition framework of attention (17).

The purpose of the present study was to explore the relevance of the biased-competition framework at the level of multivoxel fMRI response patterns. It is important to note that the effects of competition and attention in multivoxel representations cannot be understood by merely extrapolating from the single-neuron level, because several nonlinearities and unpredictable factors come into play when combining the BOLD responses to two stimuli. To cite just one example, a well controlled factor in single-neuron studies that is impossible to account for in fMRI is the control over what information is present in a given RF of a neuron: Unlike the case with single neurons when the experimenter can ensure that both stimuli are systematically presented within the RF, within a given fMRI voxel, not all neurons will necessarily respond to each stimulus present (some would respond to one stimulus, others to both, and yet others to none). Thus, one cannot trivially predict how multivoxel patterns combine or how attention affects the combination by simply observing the effects in single neurons. Furthermore, and perhaps most importantly, the relationship between neuronal signals and BOLD responses is far from being well understood (28), and, as is now frequently observed, the two signals might sometimes agree with each other, or just as easily diverge (29), depending on the experimental and perceptual conditions being tested.

In the present study, we thus explore biased competition at the level of multivoxel response patterns: Instead of focusing on the individual neuron as a system, its RF as input and its firing rate as output, as done in seminal studies of biased competition (18, 20, 21, 23), we consider here an entire brain region as the system of interest, the entire visual field as its input, and the large-scale multivoxel pattern as its output. Is the biased-competition

Author contributions: L.R. and N.G.K. designed research; L.R. performed research; R.V. contributed new reagents/analytic tools; L.R. and R.V. analyzed data; and L.R., N.G.K., and R.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: rufin.vanrullen@cerco.ups-tlse.fr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907330106/DCSupplemental.

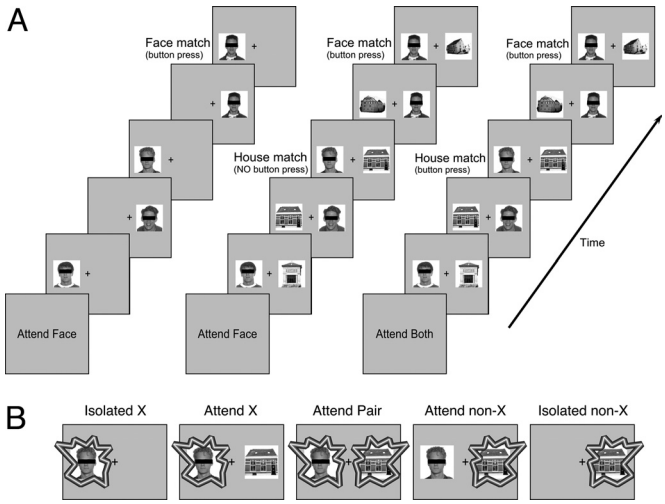

Fig. 1. Experimental design. (*A*) Subjects were presented with either one (isolated condition) (*Left*), or two (paired conditions) (*Center* and *Right*) streams of images that alternated on either side of fixation. In the paired conditions, subjects were instructed to attend to one or the other category (*Center*) or to both (*Right*). Each block began with an instruction screen presented for 2 s telling subjects which category of stimuli they were to perform a one-back task on. Images from four categories (faces, houses, shoes, and cars) were used during the experiment. (*B*) For any pair of categories X and non-X, this experimental design led to five conditions of interest. The highlighted area in each screen reflects the attended category(ies).

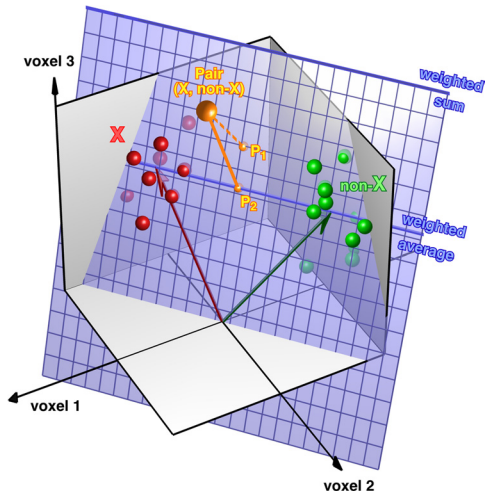

Fig. 2. A schematic representation of the analysis in a simplified 3D space, where each voxel defines a dimension. In this space, the BOLD response to different presentations of category X (presented in isolation) is shown as the cluster of red points, the response to another isolated category (non-X) is shown as the cluster of green points, and the response to the simultaneous presentation of the Pair (X, non-X) is shown as the large orange sphere. Two vectors, X and non-X (shown in red and green, respectively), represent the average position of the responses to categories X and non-X in this space. We first describe the Pair vector as a linear combination of the two vectors X and non-X by projecting it onto the plane (shown in blue) defined by these two vectors (the projection is shown as the broken orange line to point P_1). Any weighted average or weighted sum of the two vectors X and non-X also belongs to this plane, as illustrated by the two thick blue lines. By comparing the distance of the projection P_1 to the weighted average and weighted sum lines, we can determined that the Pair corresponded more closely to a weighted average response (Fig. 3; Fig. S1). Therefore, in a second step, we projected the Pair response directly onto the weighted average line (solid orange line to point P_2). The relative position of point P_2 between points X and non-X determines the weight in the weighted average response, i.e., the bias in the biased competition model. We can thus explore how this bias varies depending on the stimulus category, the region of interest, and the amount of attention directed to objects X and non-X (Figs. 4 and 5).

framework also relevant for understanding the function of such a system? To address this question we proceed in two successive and logically connected steps. First, we must determine how the response patterns corresponding to each of two simultaneously-presented stimuli combine in multidimensional voxel space. Using a novel analysis technique based on simple mathematical projections in a multidimensional space, we specifically compare predictions from two simple linear models of response combination: a weighted average or a weighted sum based on a simple linear summation of BOLD responses (30). Although there is considerable evidence for the existence of nonlinearities in neuronal responses (31, 32), this assumption of a linear combination serves here as a useful approximation because it permits expressing the paired response as the sum of two weighted components. This will conveniently allow us, in a second step, to address the main goal of this study, to characterize and quantify the effect of attention as a modification of these weights, i.e., a "bias" affecting the linear combination.

Results

Subjects were presented with stimuli from four categories (faces, houses, shoes, and cars), either in isolation or in pairs. In the latter condition, each category could be attended or unattended (i.e., attention was directed to the other category) or attention could be divided equally between the two categories. These conditions (isolated, attended, unattended, and divided attention) allowed us to look at the effects of competition and attention on large-scale multivoxel patterns of representation (Fig. 1).

Response Combination. Our general approach is illustrated in Fig. 2. The patterns of responses to each of two object categories presented in isolation (X and non-X) define a plane in the multidimensional space of possible responses (the dimensionality of the space being determined by the number of voxels in the analysis). A novel pattern recorded in response to a paired presentation of these two object categories can be projected onto this plane, i.e., it can be expressed as a linear combination of the two original patterns (plus a certain amount Δ of deviation from the plane: Pair = α.X + β.non-X + Δ). The weights α and β of this linear combination reveal the manner in which single-object responses are combined into paired responses in multidimensional voxel space: their sum $\alpha + \beta$ will be close to 1 for a weighted average combination (e.g., $\alpha = \beta = 0.5$ for the plain average) and close to 2 for a weighted sum (e.g., $\alpha = \beta = 1$ for the plain sum) (see also *SI Methods*). The actual plane projections of our data, derived from a leave-one-run-out analysis, are shown in Fig. 3, and the corresponding sums of weights in Fig. S1. Note that the projection procedure results in significant errors (the distance from the plane Δ, shown in Fig. S2), suggesting that other factors also come into play that cannot be explained by any linear combination model (measurement noise and the small number of data samples being the most obvious of these factors).

Fig. 3 illustrates the results of the plane projection separately for the four categories of objects used in this experiment (faces, houses, shoes, and cars) in three regions of interest (ROIs) [fusiform face area (FFA), parahippocampal place area (PPA), and the set of object responsive voxels in occipito-temporal cortex (ORX); see *Methods*]. In this figure, the axes have been normalized so that, in the ideal case, the isolated conditions (i.e., vectors X and non-X) would project onto the cardinal points (0,1) and (1,0), respectively. The large red and green points in Fig. 3 correspond to the two isolated presentations (e.g., in the

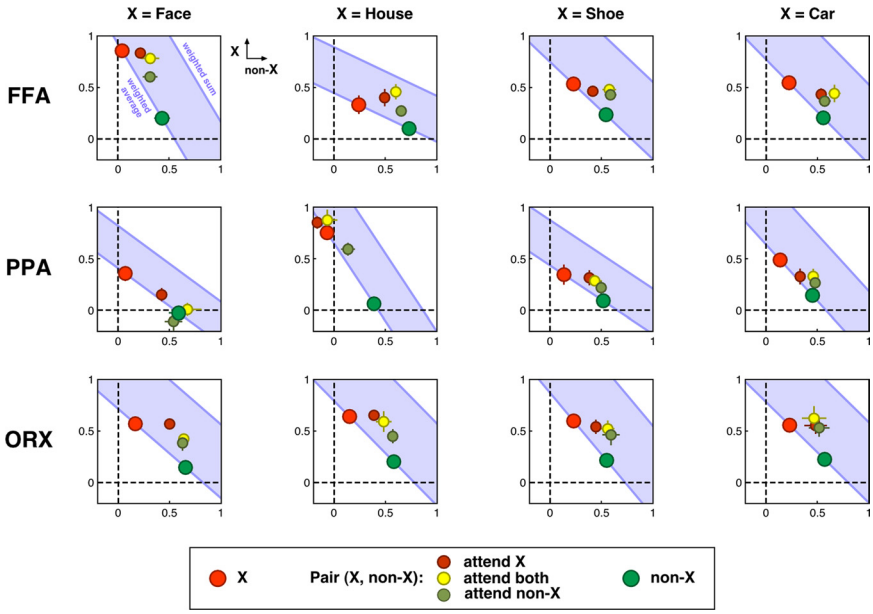

Fig. 3. The response in each condition projected onto the plane defined by the X and non-X vectors for each category and ROI. The *y* and *x* axes in each subplot correspond to the X and non-X vectors in Fig. 2 (shown here as orthogonal to each other for convenience). The two solid blue lines correspond to the families of vectors defined by the weighted average and weighted sum responses (as in Fig. 2). The projections were computed using a leave-one-run-out analysis: N − 1 runs of the isolated conditions defined the X and non-X vectors as well as the corresponding plane, and the Nth run of each of the five conditions was then projected onto the plane. This procedure was first performed pair-wise for all combinations of categories (i.e., if X = face, the projection plane and projected values were computed separately for the face-house, face-shoe, and face-car pairs), and the results were then subsequently averaged in this figure. SEM values were computed across subjects, although they are too small to be observed in most cases. Note that the X and non-X points do not lie at (0,1) and (1,0), respectively, because the leave-one-run-out analysis is influenced by the small number of runs and the variability inherent to the data.

leftmost column the large red point represents the average projection of all isolated face blocks and the large green the isolated nonface blocks). These isolated conditions do not project perfectly onto the cardinal points; this deviation is because the projections were obtained using a leave-one-out analysis: N − 1 runs were used to define the plane, on which data from the Nth run was projected; this procedure was repeated N times, each time with a different Nth run. The distance from the isolated conditions to the cardinal points thus reflects the variability intrinsic to the data and the small number of runs (typically, $n = 7$) in the leave-one-run-out analysis (also apparent in Fig. S2). The two blue lines in each plot in Fig. 3 correspond to the family of vectors defined by the weighted average response model (weights verify $\beta = 1 - \alpha$) and the weighted sum response model (weights verify $\beta = 2 - \alpha$), with the intervening space (shaded in blue) spanning intermediate possibilities between these two extreme models. The smaller points represent the paired presentation conditions, with attention either directed to category X (small red points), away from category X (small green points), or equally divided between the two (small yellow points). By comparing the location of these three types of points in the plane with the corresponding weighted average and weighted sum lines, we can determine how individual responses are combined during paired presentations.

Fig. 3 reveals that, in most cases (i.e., for all but one of the 36 observations = three paired conditions * four object categories * three ROIs), the paired response lay closer to the expected weighted average than to the weighted sum responses (the exception being house/nonhouse with equally divided attention in the FFA). The distance of the paired responses from the two models was quantified in Fig. S1 as follows: As mentioned above, the left and right blue lines in Fig. 3 correspond to linear combinations of vectors X and non-X such that the sum of weights $\alpha + \beta$ is 1 and 2, respectively. Thus, for all three paired conditions, we can compute an index between 1 and 2 that gives a measure of how far these points lie from the weighted average and weighted sum conditions. This index, based on the *y*-intercept of lines passing through the points of interest, and parallel to the blue lines, was calculated thus:

(*y*-intercept of line of interest)/(*y*-intercept of weighted-average line).

The index would be 1 for an ideal weighted average and 2 for an ideal weighted sum. The index values obtained from the data in Fig. 3 are shown in Fig. S1, collapsed over object categories but separated by ROI and paired condition. All index values were closer to 1 than to 2, i.e., they leaned toward a weighted average response.

Note that in the above analyses (Fig. 3; Fig. S1), the reference points used to calculate the weighted average and weighted sum models were the actual projections of the isolated conditions (i.e., from the left-out run in the leave-one-out analysis), rather than the cardinal points (that were determined from the remaining N − 1 runs). This makes sense, because one would expect the two isolated conditions to lie directly on their weighted average line. However, we also reach a similar conclusion (i.e., consistent with the weighted average model) by directly projecting the multidimensional paired presentation vectors (of the Nth run) onto the average and sum lines (determined from the N − 1 runs), and then comparing the corresponding projection errors (i.e., comparing Fig. S3 with Fig. S4). In fact, in this analysis, there was not a single case among the 36 data points for which the projection onto the sum line was closer than onto the average line.

To summarize, when two objects are simultaneously presented, the large-scale multivoxel patterns tend to combine in a manner more compatible with the weighted average model than the weighted sum model (even though strong departures exist; see Figs. S2–S4). The weighted average line thus provides us with a convenient axis on which to project the paired responses in multidimensional voxel space. We can now ask how attention

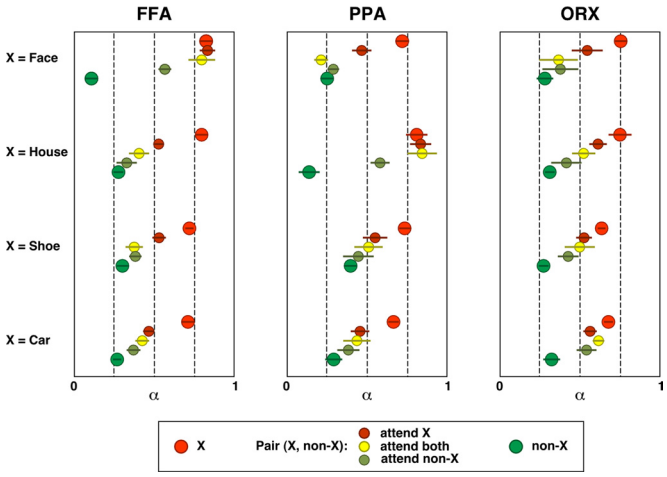

Fig. 4. Vector projections of each condition onto the weighted average line. As in Fig. 3, the projections were obtained using a leave-one-run-out analysis, N − 1 runs of the two isolated conditions defined the weighted average line, and the Nth run was then projected onto the line. For each category X, the procedure was repeated pair-wise for all three combinations of (X, non-X), and the averaged results are shown here (as in Fig. 3). SEM values were calculated across subjects. The points for each X/non-X are staggered along the *y* axis to avoid superposition. The *x* axis represents the value (between 0 and 1) of the weight α in the weighted average: α.X + (1 − α).non-X. The effects of category and attention on this bias can be viewed as a ''slider'' that shifts the paired representation along the axis joining the two individual representations.

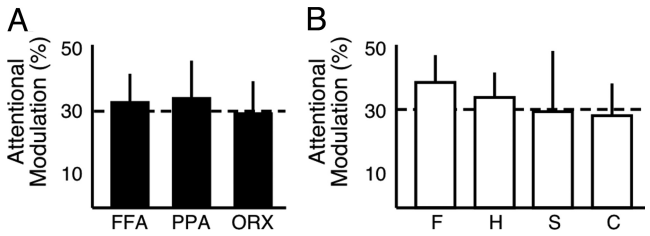

Fig. 5. Attention modulates the weights by ≈30% in each ROI (*A*; collapsed across categories) and for each category (*B*; collapsed across ROIs). The strength of the attentional modulation was computed as the amount (in Fig. 4) by which the weight of the ''attend-both'' condition shifted toward the ''isolated X'' or ''isolated non-X'' weight in the ''attend-X'' and ''attend-non-X'' conditions, respectively.

alters the position of the response pattern along this axis, i.e., how attention biases the competition between simultaneously-presented stimuli.

Attention and Biased Competition. Having established that multivoxel fMRI patterns combine more in line with a weighted average model, we now turn to the main goal of this study and examine the specific influence of attention on paired responses. More precisely, we ask how attention biases the weights of the weighted average response. To this end, in a leave-one-run-out analysis, we projected each paired condition of the Nth run onto the line joining the two vectors defined by the isolated conditions (in the N − 1 runs) of the two categories of interest. This projection corresponds to the point marked P_2 in Fig. 2. The resulting projection can be expressed as a weighted average (with weight α) of the two isolated vectors (plus a certain amount Δ' of deviation from the line: Pair = α.X + (1 − α).non-X + Δ'). Note that the projection procedure again results in important errors (the distance from the line Δ', shown in Fig. S3), highlighting the variability, the small number of runs, and/or the nonlinearity of our fMRI dataset. However, applying the same procedure to the weighted sum model yields projections (satisfying: Pair = α.X + (2 − α).non-X + Δ'') that are even more distant from the original data (as can be seen in Fig. S4, illustrating the distance Δ''). This result confirms our choice of the weighted average as the optimal linear combination model.

Fig. 4 reports the values of the combination weight α for the four object categories in the three ROIs, as a function of the attentional condition: category X or non-X presented in isolation (large red and green points, respectively), paired presentation with attention directed to category X or non-X (small red and green points, respectively), or divided equally between both (small yellow points). As in the previous analysis, the projections for the two isolated blocks do not lie at their ideal position (i.e., α = 0 and α = 1), due to the leave-one-out analysis and the variability in our dataset. However, the fact that these two points are well separated in each ROI and for all categories indicates that meaningful category information can be extracted from these multivoxel patterns, as already demonstrated by several previous studies (1, 2, 4, 5). The weight value when attention is equally divided between two simultaneously-presented categories (represented by the position of the yellow points in Fig. 4) reflects a large category bias in the FFA and PPA for faces and houses, respectively. In these two areas, the paired presentation is almost superimposed onto the isolated presentation of the ''preferred'' stimulus category, meaning that the weight of the preferred category is ≈1 (when it is expressed relative to the isolated conditions of the left-out run in the leave-one-out analysis, i.e., the large red and green points, rather than the absolute points of 0 and 1 determined from the N − 1 runs). In other words, in these regions, the preferred stimulus acts as an attractor for the paired response. In ORX, however, for all categories, the divided attention condition lies more or less halfway between the two isolated conditions, i.e., with a weight of ≈0.5. Finally, the attended and unattended (albeit to a lesser extent) conditions also tend to follow the category bias. These findings of an influence of stimulus category and ROI on the robustness of dual-image representations (statistically confirmed by the ANOVA described below) are compatible with a previous report (33).

To statistically evaluate the effect of category and attention on the magnitude of the weights, we performed a three-way ANOVA of attention (attended, divided, and unattended) × category × ROI with the weight as the dependent variable. We observed a main effect of attention [$F(2,324) = 16.04$; $P < 0.0001$] and a posthoc analysis revealed that the weights were ordered by attended > divided > unattended (corrected for multiple comparisons with Scheffé's method), as can easily be observed in Fig. 4. We also observed a main effect of category [$F(3,324) = 3.3$; $P = 0.02$] and an interaction effect of category × ROI [$F(6,324) = 21.46$; $P < 0.0001$] as expected from the category preferences of the FFA and PPA. No two- or three-way significant interaction effects were observed with attention, suggesting that the attentional effect is largely independent of both category and ROI.

Finally, to investigate how attention modifies the weights, we computed the amount by which attention causes the weight in the divided attention condition to shift toward either attended category (Fig. 5; Fig. S5). This measure is, basically, the bias in the biased-competition theory. We found that there was approximately a 30% shift in the weights with attention, quantitatively consistent with neurophysiological results from the monkey literature (21, 23–27). A two-way ANOVA of category × ROI with this attentional bias as the dependent variable did not reveal a significant main effect of category [$F(3,104) = 0.27$; $P = 0.85$] or ROI [$F(2,104) = 0.1$; $P = 0.9$] or any interaction effect [$F(6,104) = 0.2$; $P = 0.98$]. Thus, the 30% attentional shift was constant, regardless of category or ROI (Fig. 5).

Discussion

This study tested the relevance of the biased-competition framework at the level of multivoxel patterns of BOLD activity. We considered entire brain regions as our system of interest, the entire visual field as its input, and the multidimensional pattern of voxel responses as its output, and asked whether competition takes place between simultaneous stimuli, and how attention modulates this competition. We found that the response to a pair of stimuli could be approximated by the average of their individual responses in a multidimensional voxel space (Fig. 3; Fig. S1) (34); preferred categories in selective regions tended to act as attractors for the paired response (Fig. 4) (33); finally, attention biased the paired response by ≈30% (Fig. 5). Despite the fundamental differences between neuronal firing rates and BOLD signals discussed in the introduction, these results are surprisingly consistent with a large body of electrophysiological literature showing that the neuronal response to a pair of visual stimuli can be well described as an average of the response to the two constituent stimuli (21), even though evidence for nonlinear components also exists (31, 32), and that attention biases the averaged response in favor of the attended stimulus (18–21, 24). Furthermore, the degree of the attentional bias that we found in multivoxel patterns was quantitatively comparable with that observed in single neurons in areas V4 (21, 27) and MT (25). Above and beyond confirming single-neuron studies at the level of fMRI multivariate patterns, these results demonstrate that the biased-competition framework is also relevant at the level of large-scale, integrated perceptual representations.

Our assessment of response combination as a weighted average rather than a weighted sum was intended as a first linear approximation (and in reality, the paired responses generally fall in between these two linear predictions) (Fig. 3). Recent studies tend to support more elaborate, nonlinear pooling schemes (21, 31, 32, 35). In particular, the normalization model of attention (a powerful implementation of the biased-competition model using a normalization factor and taking into account differences in the sensory input to the neurons) (35, 36) can account for various attentional effects such as response gain (37, 38), contrast gain (39, 40), or sharpening of the neuronal tuning curves (41, 42). The normalization model predicts that, under sensory conditions that drive neurons to their saturation point (i.e., at high contrast values), the paired response is consistent with a weighted average model; however, at lower contrasts, the combined response is expected to be larger than an averaged response. This prediction of an intermediate paired response between a weighted average and weighted sum is actually consistent with our findings (Fig. 3), and suggests that the normalization model of attention may also be relevant in the domain of multivoxel patterns. Further work will be needed, however, to systematically extend our multivoxel approach to this type of nonlinear framework.

Attentional effects for multiple simultaneous stimuli have previously been explored in fMRI studies with a focus on the univariate response of voxels or ROIs. Kastner et al. (43) showed that the average BOLD response over a given ROI to simultaneous stimuli was smaller than the response to a sequential presentation of the same stimuli, but that attention counteracted this suppression. Bles et al. (44) reported that the spatial extent of attentional modulation (as measured by the average BOLD) was correlated with estimated RF size across visual cortex. However, as is becoming increasingly evident, the univariate BOLD response will often fail to provide information that can otherwise be gleaned from multivariate patterns of BOLD activity (1, 4, 5, 11). For instance, in our experiment, the average BOLD response in ORX to isolated presentations of the four object categories is approximately equivalent (Fig. S6), which would lead one to conclude that this region carries no discriminatory information about object category, an obviously misguided conclusion, as demonstrated by the separation between isolated conditions in Fig. 4 *Right*. Given that the subjects in our experiment could easily discriminate between these categories, in this instance, the multivariate patterns turn out to be more representative of subjects' perception than the univariate BOLD response (although this is not always the case) (45). Therefore, the question of attention and biased competition must also be addressed with multivariate fMRI patterns, which was the purpose of the present study.

Decoding the attentional state from multivoxel fMRI patterns of activity was discussed recently in a study with a focus on earlier visual areas (4). The authors demonstrated that multivoxel patterns in V1 and V2 allow an ideal observer to determine which of two simultaneously-presented orientations is currently attended (4). However, neural representations in early visual areas are local and retinotopic, and it is not clear if the same principles would extend to higher-level visual areas of ventral cortex with less retinotopic representations and more global RFs; in fact, the above-mentioned finding did not generalize easily to areas V3 and V3A/V4 (4, 46). One previous study has considered the effects of competition between simultaneous stimuli and attention on multivoxel ventral-temporal cortex representations, setting the stage for the current work (33). The purpose of that decoding study was to evaluate the ability of a given ROI to signal the presence of its preferred category regardless of attentional state, in other words manipulations of attention were regarded as a potential source of confusion or "noise" for the decoding task being tested. In contrast, here we evaluate how strongly attention "slides" the multivoxel response to any pair of stimuli toward the response vector of the attended category (regardless of whether it is the preferred category or not for the ROI under consideration). In other words, in the present study, attention served to disambiguate the overlapping multidimensional patterns caused by paired stimulus presentations, thereby increasing object discriminability. This finding is consistent with the predictions of the biased competition framework, and demonstrates the relevance of this framework for large scale multivariate representations.

Methods

Subjects. Ten healthy subjects participated in the fMRI study. All subjects gave signed consent and had normal or corrected-to-normal vision.

Experimental Timeline. Stimuli were presented with Psychophysics Toolbox. In separate blocks, images from four categories of objects (faces, houses, shoes, and cars) were presented either in isolation, or in pairs. In the latter "pair" condition, subjects were instructed to attend to one or the other image category or simultaneously to both categories. As shown in Fig. 1, this design allowed for five attentional conditions for each category. To ensure that subjects attended to the category in question, they were required to perform a one-back task on the images. In each block of 20 stimulus presentations, a one-back repeat occurred twice per category, presented at random times. Each block was 16-s long and each stimulus was presented for 800 ms. The stimuli and paradigm were essentially identical to those used in ref. 33, with the addition of the dual task condition that was essential for quantifying response combination and attention effects. For further details, see ref. 33 and *SI Methods*.

ROIs. In separate localizer runs, subjects were presented with blocks of faces, scenes, objects, and scrambled images. Based on the data obtained in these localizer runs, three ROIs were defined. The FFA was defined as the set of contiguous voxels in the fusiform gyrus that showed significantly stronger activation ($P < 10^{-4}$, uncorrected) to faces than to other objects. The PPA was defined as the set of voxels in the parahippocampal gyrus that showed stronger activation to scenes versus objects ($P < 10^{-4}$, uncorrected). The ORX ROI was the set of distributed voxels in the ventral temporal cortex that were more strongly activated to objects in general, with no preferred selectivity for any particular category. This ROI was defined by a contrast of faces, objects, or scenes versus scrambled images ($P < 10^{-5}$, uncorrected), with the exclusion of

any voxels that were specifically selective for faces and scenes. Note that ORX does not simply correspond to LOC (that is generally defined as the set of voxels in the lateral occipital and posterior fusiform regions that respond more strongly to objects versus scrambled objects), but also includes other object responsive voxels distributed in ventral temporal cortex.

Functional MRI Data Acquisition and Analysis. Functional MRI data were collected on a 3T Siemens scanner (gradient echo pulse sequence, TR = 2 s, TE = 30 ms, 20 slices with a 12 channel head coil, slice thickness = 2 mm, in-plane voxel dimensions = 1.6 × 1.6 mm). The slices were positioned to cover the entire temporal lobe and part of the occipital lobe. Data analysis was performed with FreeSurfer Functional Analysis Stream (FS-FAST) (http://surfer.nmr.mgh.harvard.edu), fROI (http://froi.sourceforge.net), and custom Matlab scripts.

ACKNOWLEDGMENTS. We thank Francisco Pereira for valuable comments on the manuscript. This work was supported by grants from the Fondation pour la Recherche Médicale and the Fyssen Foundation (L.R.); the National Eye Institute Grant EY 13455 (to N.G.K.); and the Agence Nationale de Recherches Project ANR 06JCJC-0154 and the Fyssen Foundation and the European Young Investigator Award (R.V.).

1. Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
2. Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representations of objects. *J Cognitiv Neurosci* 15:704–717.
3. Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) ''brain reading'': Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19:261–270.
4. Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685.
5. Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8:686–691.
6. Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452:352–355.
7. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868.
8. Haynes JD, Rees G (2005) Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol* 15:1301–1307.
9. Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–635.
10. Stokes M, Thompson R, Cusack R, Duncan J (2009) Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J Neurosci* 29:1565–1572.
11. Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
12. Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev* 7:523–534.
13. Corbetta M, et al. (1998) A common network of functional areas for attention and eye movements. *Neuron* 21:761–773.
14. Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22:751–761.
15. McMains SA, Somers DC (2004) Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42:677–686.
16. Serences JT, Boynton GM (2007) Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55:301–312.
17. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222.
18. Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
19. Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol* 70:909–919.
20. Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77:24–42.
21. Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753.
22. Reynolds JH, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24:19–29:111–125.
23. Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382:539–541.
24. Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. *Neuron* 37:853–863.
25. Treue S, Maunsell JH (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci* 19:7591–7602.
26. Deco G, Rolls ET (2005) Neurodynamics of biased competition and cooperation for attention: A model with spiking neurons. *J Neurophysi* 94:295–313.
27. Fallah M, Stoner GR, Reynolds JH (2007) Stimulus-specific competitive selection in macaque extrastriate visual area V4. *Proc Natl Acad Sci USA* 104:4165–4169.
28. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
29. Maier A, et al. (2008) Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat Neurosci* 11:1193–1200.
30. Hansen KA, David SV, Gallant JL (2004) Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *NeuroImage* 23:233–241.
31. Heuer HW, Britten KH (2002) Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysi* 88:3398–3408.
32. Gawne TJ, Martin JM (2002) Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysi* 88:1128–1135.
33. Reddy L, Kanwisher N (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol* 17:2067–2072.
34. Macevoy SP, Epstein RA (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol* 19:943–947.
35. Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
36. Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Visual Neurosci* 9:181–197.
37. McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441.
38. Treue S, Martinez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579.
39. Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26:703–714.
40. Martinez-Trujillo J, Treue S (2002) Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron* 35:365–370.
41. Spitzer H, Desimone R, Moran J (1988) Increased attention enhances both behavioral and neuronal performance. *Science* 240:338–340.
42. Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14:744–751.
43. Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282:108–111.
44. Bles M, Schwarzbach J, De Weerd P, Goebel R, Jansma BM (2006) Receptive field size-dependent attention effects in simultaneously presented stimulus displays. *NeuroImage* 30:506–511.
45. Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685–686.
46. Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage* 44:223–231.

Supporting Information

Reddy et al. 10.1073/pnas.0907330106

SI Methods

Experimental Design. As shown in Fig. 1, within each block, the images of a given category were alternately presented to the left and right of fixation. This design prevented subjects from attending to just one side of fixation during an entire block, thus avoiding potential confounds due to hemisphere-specific activations. Although our attentional manipulation clearly entails a "category-specific" component (only one or two object categories were task-relevant at any time), it may also involve spatial attention, because the 800-ms stimulus duration and the predictability of the alternating sequence would have allowed subjects to spatially attend to the relevant location on each stimulus presentation. The images were centered at 4° on either side of fixation and subtended ≈7 × 7° of visual angle. Each subject was tested on six to seven experimental runs in the scanner. The order of blocks was counterbalanced across subjects. Eye tracking was performed in a separate session with an IR IScan Camera, sampling the eye position at 240 Hz. Viewing conditions in the scanner were replicated as closely as possible during these sessions. Subjects were not explicitly instructed to fixate during these sessions but were told to perform the experiment with the same strategy they had used during the fMRI scanning session (when they had been strictly instructed to fixate). As reported in detail in ref. 1, subjects reliably maintained fixation and did not make eye movements to the peripherally presented stimuli.

Projection Analysis. As also described below (see also Fig. 2), our analysis relied on simple mathematical projections in a multidimensional space: Each fMRI activation pattern was associated with a vector in multidimensional space, which was then projected either onto a plane (Fig. 3; Figs. S1 and S2) or onto a line (Fig. 4; Figs. S3 and S4). The plane and line were defined (as described in more detail below) based on two reference vectors, corresponding to isolated presentations of categories X and non-X. For both the plane and line analyses, a leave-one-out procedure was used: The isolated conditions in N − 1 runs were used to define the plane or the weighted average line, respectively, and data from the Nth run was then projected onto this plane or weighted average line. Additionally, for each category X, all analyses were performed pair-wise (e.g., face-house, face-shoe, and face-car) and then averaged to give a single value for (X, non-X).

Plane Projections. The first step of the analysis asked whether the joint response during paired presentation, Pair (X, non-X), corresponds better to a weighted average of the response to the individual stimuli (X, non-X) or to a weighted sum.

That is, if Pair (X, non-X) = α.X + β.non-X + Δ, is $\alpha+\beta \sim 1$ (the weighted average model), or is $\alpha + \beta \sim 2$ (the weighted sum model)?

The responses to X and non-X in the multidimensional voxel space correspond to high-dimensional vectors, and the plane containing these two vectors also contains the weighted average and weighted sum lines (see Fig. 2). Thus, to determine which model better describes the response in the paired presentation condition, the corresponding vector Pair (X, non-X) can be projected onto this plane. The distance of this projected point from the weighted average and weighted sum lines indicates which model better suits the data; the shorter the distance, the better the model.

The projection of a point P to point P_1 on the plane containing vectors X and non-X can be computed by solving the following three equations:

$$\vec{P}_1 = \alpha\cdot\vec{X} + \beta\cdot\overrightarrow{\text{non-X}} \quad [1]$$

$$(\vec{P}_1 - \vec{P})\cdot\vec{X} = 0 \quad [2]$$

$$(\vec{P}_1 - \vec{P})\cdot\overrightarrow{\text{non-X}} = 0 \quad [3]$$

Solving for α and β yields:

$$\alpha = \frac{(\vec{P}\cdot\overrightarrow{\text{non-X}})\cdot(\vec{X}\cdot\overrightarrow{\text{non-X}}) - (\vec{P}\cdot\vec{X})\cdot|\overrightarrow{\text{non-X}}|^2}{(\vec{X}\cdot\overrightarrow{\text{non-X}})\cdot(\vec{X}\cdot\overrightarrow{\text{non-X}}) - |\vec{X}|^2\cdot|\overrightarrow{\text{non-X}}|^2}$$

$$\beta = (\vec{P}\cdot\vec{X} - \alpha\vec{X}\cdot\vec{X})/(\vec{X}\cdot\overrightarrow{\text{non-X}})$$

Note that $|\vec{P}_1 - \vec{P}| = |\vec{\Delta}|$ would correspond to the distance of P from the plane, and would thus be a measure of how much of the response to Pair (X, non-X) is not explained by a simple linear model (see Fig. S2).

Weighted Average Line Projections. The distance of the plane projection P_1 of Pair (X, non-X) from the weighted average and weighted sum lines indicated that the projection error was lower for the weighted average model (Fig. 3; Fig. S1). In the second step of the analysis, we therefore asked, for all paired responses, what values of α and β (i.e., the weights) satisfied the equation: Pair (X, non-X) = α.X + β.non-X + Δ', such that $\alpha+\beta = 1$ for all of the paired responses. Note that the constraint $\alpha+\beta = 1$ implies that the projection $P_2 = \alpha$.X + β.non-X lies along the line joining points X and non-X (i.e., the weighted average line).

To solve this equation for α (and $\beta = 1 - \alpha$), we thus projected Pair (X, non-X) onto the line between points X and non-X, using the following equation:

$$\alpha = (\vec{X} - \overrightarrow{\text{non-X}})\cdot(\vec{P} - \overrightarrow{\text{non-X}})/|\vec{X} - \overrightarrow{\text{non-X}}|^2$$

The value $|\vec{P}_2 - \vec{P}| = |\vec{\Delta}'|$ in this case reflects the variability that the weighted average model cannot account for (Fig. S3). Note that some authors have also argued in favor of a "max" model (2, 3), the projection error for the max model was also computed but turned out to be larger than for the weighted average model, and was thus not further considered in the analysis.

The attentional bias was computed as the ratio of the distance between the "attend-X" (or "attend-nonX") point in Fig. 4 to the "isolated X" (or "isolated non-X") point, and the distance between the "attend both" point to the "isolated-X" (or "isolated-nonX") point. To avoid dividing by zero or by very small numbers (that would artificially inflate the index in either the positive or negative directions), we did not consider instances where the "attend-both" condition was implausibly close to either isolated condition (i.e., if the denominator in the ratio was <0.1).

1. Reddy L, Kanwisher N (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol* 17:2067–2072.
2. Gawne TJ, Martin JM (2002) Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysiol* 88:1128–1135.
3. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.

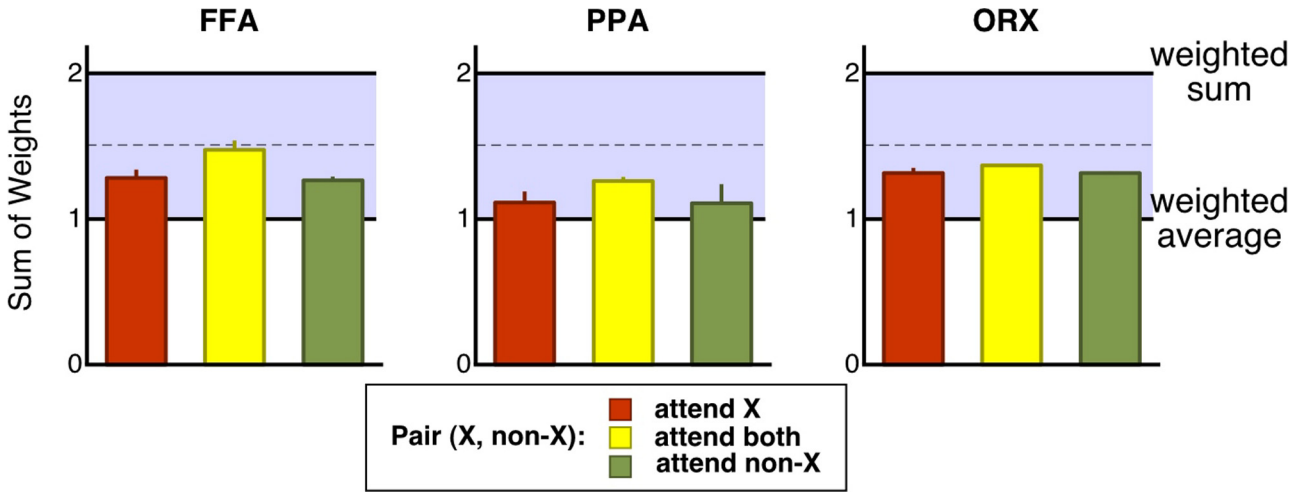

Fig. S1. An index evaluating the distance of each Pair (X, non-X) in Fig. 3 to the expected weighted average and weighted sum lines for the fusiform face area (FFA), parahippocampal place area (PPA), and object responsive voxels in ventral temporal cortex (ORX); see *Methods*. A value of 1 would reflect a true weighted average, and a value of 2 a true weighted sum. The shaded blue area corresponds to the blue area in Fig. 3.

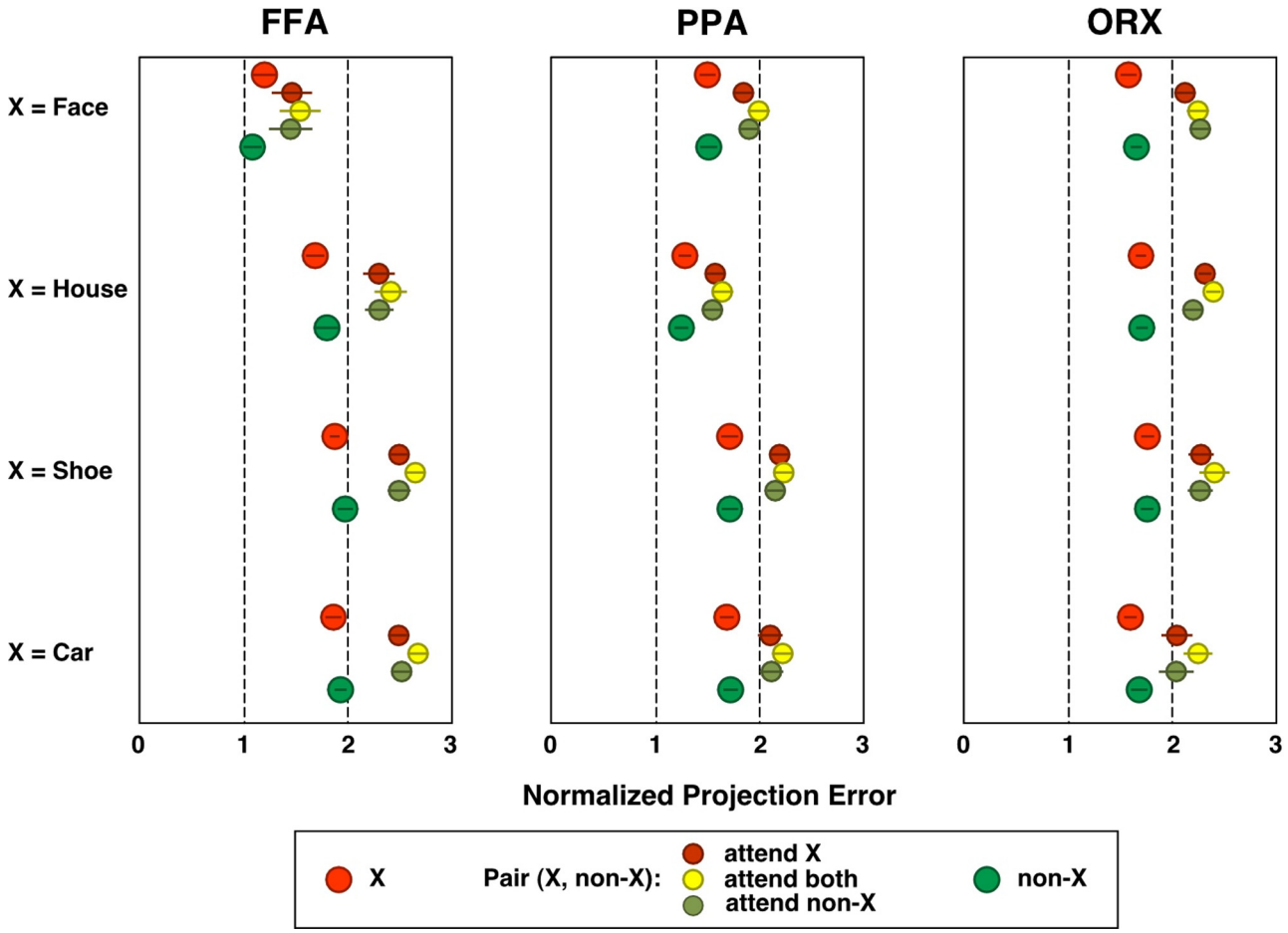

Fig. S2. The normalized projection error quantifying the distance of the response in each condition from the plane defined by the X and non-X vectors (normalized to the norm of vector [X, non-X]). As in Figs. 3 and 4, for each X, this analysis was computed pair-wise over all combinations of (X, non-X) and the averaged results are shown here. SEM values were calculated across subjects. The normalized projection errors for the isolated conditions reflect the block-to-block variability in the dataset (i.e., the responses in different repetitions of a given condition are not identical). The additional normalized projection error values for the pairs (X, non-X), beyond those observed for isolated presentations reflect the variability (and noise) that neither the weighted sum nor weighted average models can account for. The projection error for the paired conditions (X, non-X) is only 34 $\pm$ 2% larger than the corresponding projection error for the isolated conditions. Therefore, the inherent variability and measurement noise that similarly affect the isolated and paired conditions account for the larger portion of the projection error.

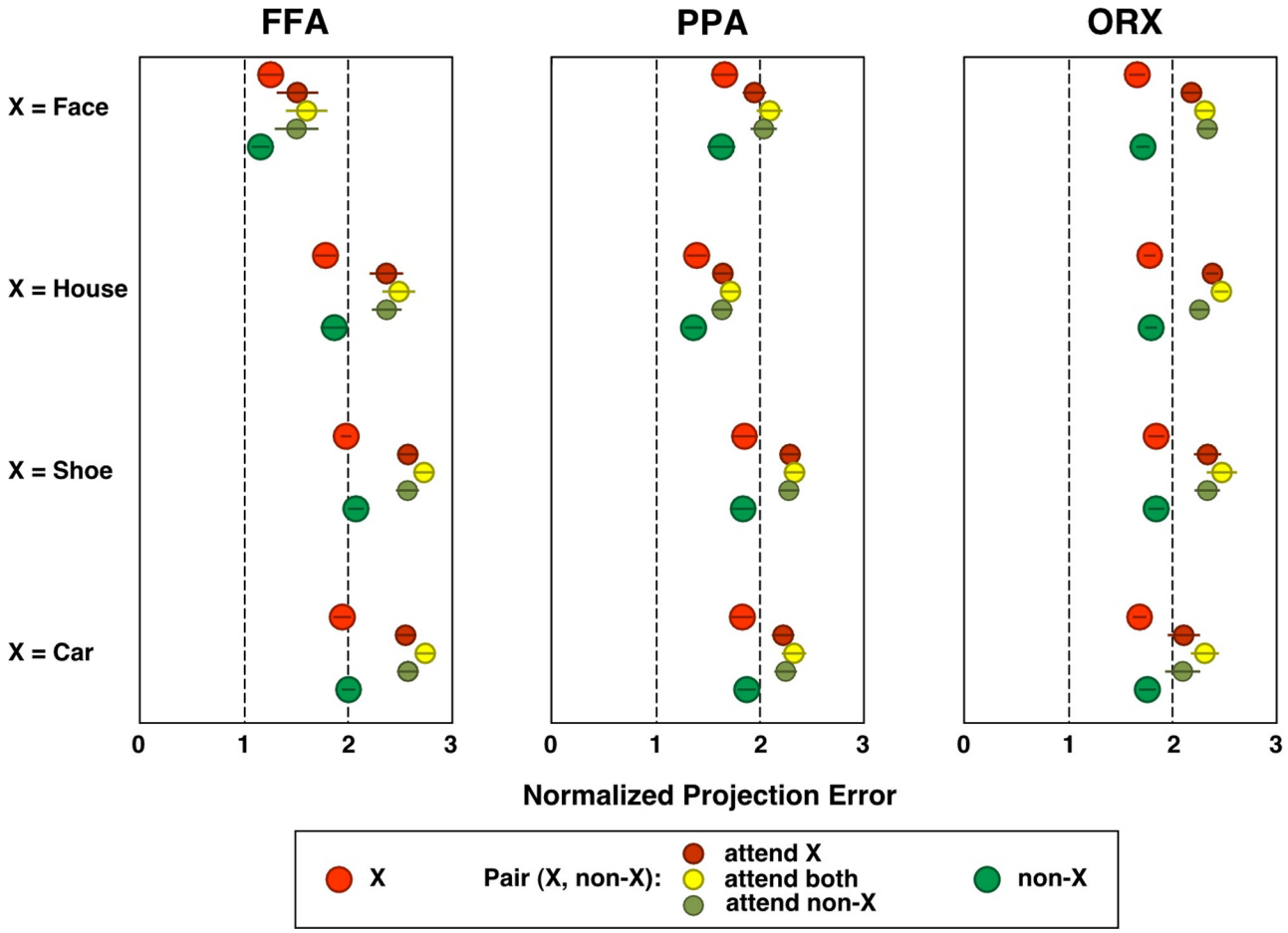

Fig. S3. The normalized projection error quantifying the distance of the response in each condition from the weighted average line defined by the X and non-X vectors (normalized to the distance between points X and non-X). As previously, for each X, this analysis was computed pair-wise over all combinations of (X, non-X) and the averaged results are shown here. SEM values were calculated across subjects. The projection error for the paired conditions (X, non-X) is only 31 ± 2% larger than the corresponding projection error for the isolated conditions. Note that even though Figs. S2 and S3 look quite similar to each other slight differences can be observed. The average normalized projection error is 1.95 in Fig. S2 and 2.02 in Fig. S3. Furthermore, the projection error for each point in Fig. S3 is larger than the error for the corresponding point in Fig. S2 as expected (because the line belongs to the plane, plane projection errors cannot be larger than line projection errors).

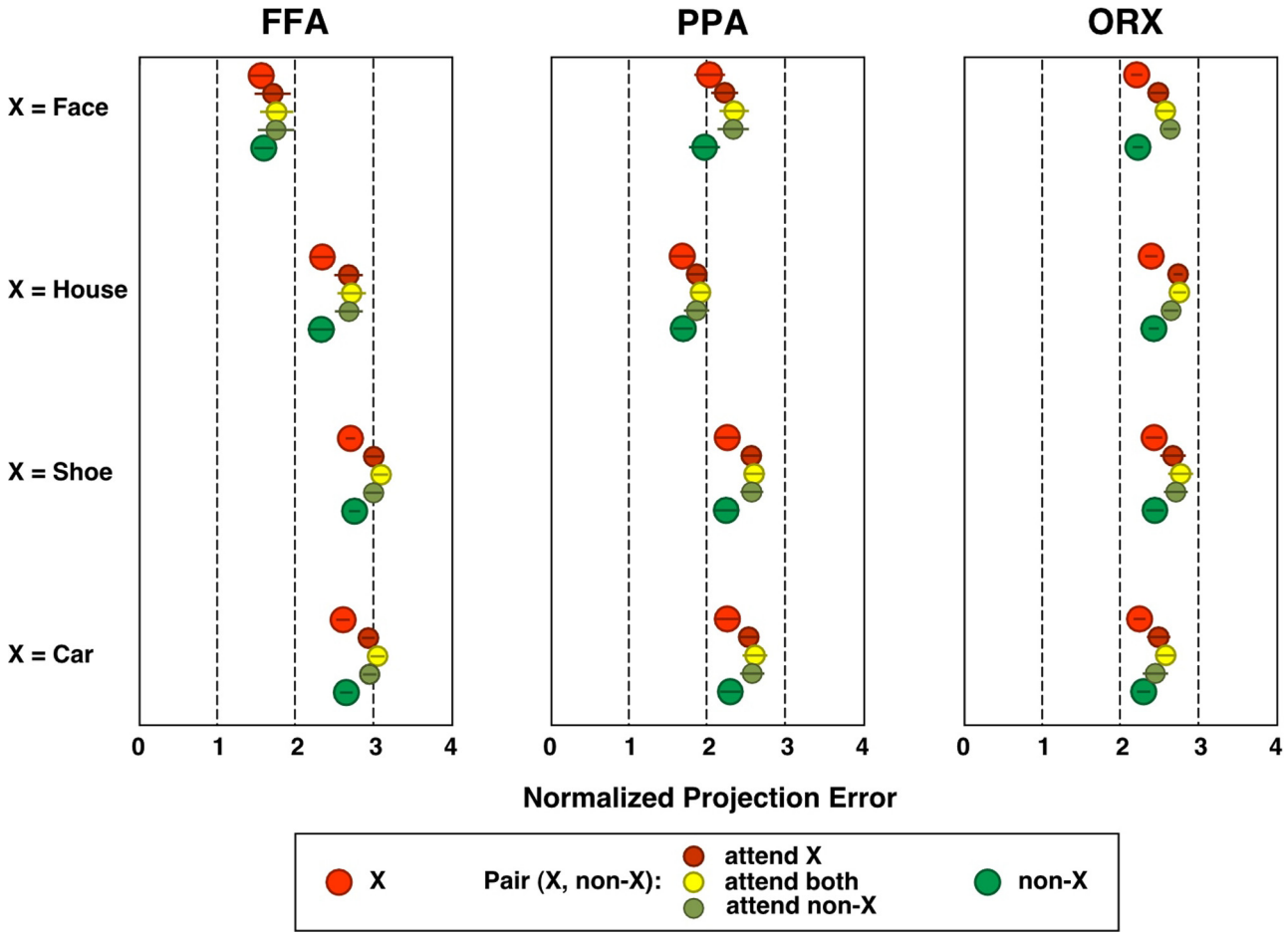

Fig. S4. The normalized projection error quantifying the distance of the response in each condition from the weighted sum line defined by the X and non-X vectors (normalized to the distance between points X and non-X). As previously, for each X, this analysis was computed pair-wise over all combinations of (X, non-X) and the averaged results are shown here. SEM values were calculated across subjects. The averaged normalized projection error from the weighted sum line was larger than the corresponding distance from the weighted average line (compare with Fig. S3; note the difference in the *x*-axis scale).

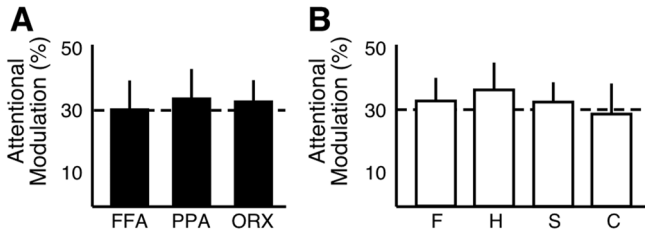

Fig. S5. In a second measure of the attentional bias, we considered the bias to be the distance between the attended conditions normalized by the distance between the isolated conditions on the weighted average line projection. This measure has the advantage of depending only on the projections of the attend-X and attend-nonX conditions (and of the corresponding isolated conditions), but not on the attend-both condition. The results of this analysis are consistent with the attentional bias reported in Fig. 5.

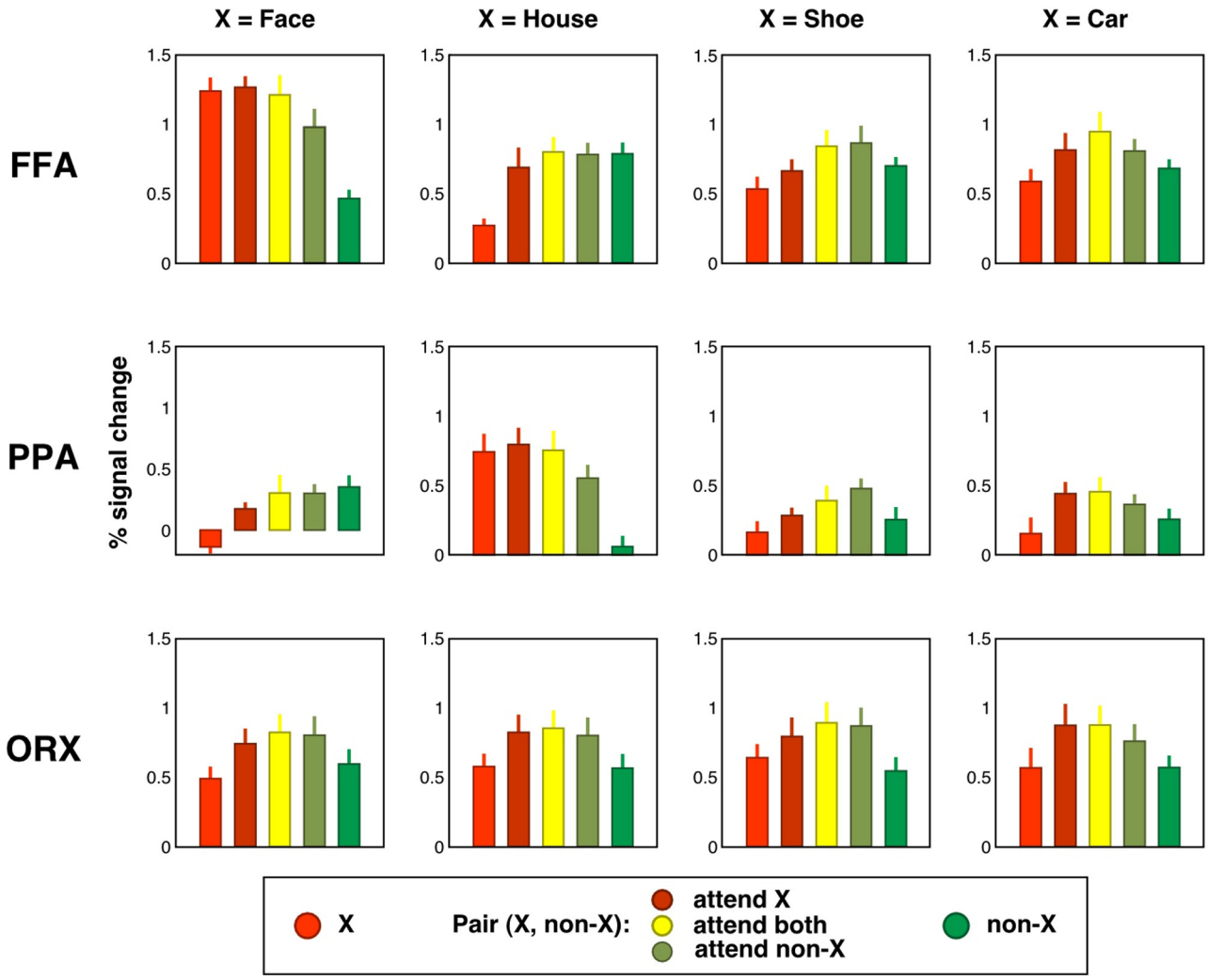

Fig. S6. The blood-oxygen-level-dependent (BOLD) signal change in each region of interest (ROI) for the four categories in the five experimental conditions. Note that the absolute level of BOLD response for isolated stimulus presentations sometimes does not allow one to distinguish between categories (e.g., for all categories in ORX), whereas the multivoxel response does (see Fig. 4). Furthermore, the absolute BOLD response does not reveal a consistent pattern for how multiple stimuli are combined. For instance, in those cases where a ''preferred'' and a ''nonpreferred'' category exist (e.g., for faces and houses in the FFA and PPA), the response to paired presentations tends to fall between the isolated responses (as predicted by a weighted average model). However, when no preferred stimulus is apparent (e.g., shoes and cars in all three ROIs), there appears to be an additive component in the paired response (suggestive of a weighted sum model). In contrast, looking at response combination in the multivoxel space did reveal a consistent pattern, across categories and ROIs, in favor of a weighted average model (Fig. 3; Fig. S1).