Non-holistic coding of objects in lateral occipital complex with and without attention

Matthias Guggenmos[1,2], Volker Thoma[3], Radoslaw Martin Cichy[4], John-Dylan Haynes[1], Philipp Sterzer[1,2†] & Alan Richardson-Klavehn[5,6†]

[1] Bernstein Center for Computational Neuroscience, Berlin, Germany
[2] Visual Perception Laboratory, Charité Universitätsmedizin, Berlin, Germany
[3] School of Psychology, University of East London, London, UK
[4] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, USA
[5] Department of Neurology, Otto von Guericke University, Magdeburg, Germany
[6] Leibniz Institute for Neurobiology, Magdeburg, Germany

[†]contributed equally

Corresponding author:
Matthias Guggenmos
Bernstein Center for Computational Neuroscience, Philippstraße 13, Haus 6, 10115 Berlin
Phone: +49 (0) 30 450 517131
E-Mail: matthias.guggenmos@bccn-berlin.de

Abstract

A fundamental issue in visual cognition is whether high-level visual areas code objects in a part-based or a view-based (holistic) format. By examining the viewpoint invariance of object recognition, previous behavioral and neuroimaging studies have yielded ambiguous results, supporting both types of representational formats. A critical factor distinguishing the two formats could be the availability of attentional resources, as a number of studies have found greater viewpoint invariance for attended compared to unattended objects. It has therefore been suggested that attention is necessary to enable part-based representations, whereas holistic representations are automatically activated irrespective of attention. In this functional magnetic resonance imaging study we used a multivariate approach to probe the format of object representations in human lateral occipital complex (LOC) and its dependence on attention. We presented human participants with intact and half-split versions of objects that were either attended or unattended. Cross-classifying between intact and split objects, we found that the object-related information coded in activation patterns of intact objects is fully preserved in the patterns of split objects and vice versa. Importantly, the generalization between intact and split objects did not depend on attention. Our findings demonstrate that LOC codes objects in a non-holistic format, both in the presence and absence of attention.

Keywords: object recognition, attention, lateral occipital complex, multivariate pattern analysis, fMRI

1. Introduction

A hallmark of human object perception is the recognition of objects despite variations in their exact appearance. Accordingly, object representations in high-level visual brain areas have to be able to generalize across changes in size, position or orientation (Eger et al., 2008; Grill-Spector et al., 1999). Yet, the specific representational code realizing such invariant representations is still largely unknown. One central question is whether an object is coded as a collection of *parts* (Hummel and Biederman, 1992; Marr and Nishihara, 1978) or in a *view-based* format (Edelman and Bülthoff, 1992; Poggio and Edelman, 1990; Tarr and Pinker, 1989).

Part-based models propose that objects are encoded in terms of their constituent parts, the representations of which are independent of each other and dynamically bound together. Neurons that are tuned to a particular object part could therefore respond to the object part appearing in different configurations or views, allowing for robust object recognition across various manipulations, such as translation across the visual field, size changes and left-right reflection (Hummel and Biederman, 1992; Hummel, 2001). By contrast, view-based models propose that objects are recognized by matching the incoming sensory information to stored views (Edelman and Bülthoff, 1992; Poggio and Edelman, 1990; Tarr and Pinker, 1989). View-based representations are holistic, as the parts of an object are not represented independent of each other and have fixed relative positions (static binding). Under a view-based scheme neurons respond most strongly if objects are presented in learned views or configurations. Nevertheless, recognition of objects in varying orientations is thought possible by storing many views of an object (Bülthoff and Edelman, 1992; Olshausen et al.,

1993; Poggio and Edelman, 1990; Tarr and Gauthier, 1998; Tarr, 1995; Ullman, 1998), interpolating across these views (Logothetis et al., 1994; Poggio and Edelman, 1990; Ullman, 1989) or by a distributed neural representation across view-tuned neurons (Perrett et al., 1998).

Behavioral evidence on the format of object representations is equivocal, supporting both view-based (Edelman and Bülthoff, 1992; Murray, 1999; Tarr and Pinker, 1989) and part-based representations (Biederman and Cooper, 1991; Biederman and Gerhardstein, 1993). Neuroimaging research, too, has sought to establish which format of representation underlies object recognition. Studies using functional magnetic resonance imaging (fMRI) show that blood oxygen level-dependent (BOLD) signals in various ventral visual stream regions, such as in lateral occipital and inferior temporal cortices, tend to decrease when an object is shown repeatedly and found that this repetition suppression (Grill-Spector et al., 2006) is greatest when the repeated view of an object is identical to the original orientation, but decreases with the amount of view change (Andresen et al., 2009; Ewbank et al., 2005; Gauthier et al., 2002). However, in support for part-based representations, other fMRI studies have shown that the ventral stream is largely insensitive to the deletion of local image features or changes in image format (grayscale image vs. line drawing), as long as the individual object parts are present (Hayworth and Biederman, 2006; Kourtzi and Kanwisher, 2000).

Importantly, the representational format might be dependent on the absence or presence of attention. Attended visual objects exhibit robust repetition-priming effects even when their mirror-reflected (Stankiewicz et al., 1998) or half-split (Thoma and Henson, 2011; Thoma et al., 2004) versions are presented as prime stimuli, suggesting a part-based

representation. However, when the same prime objects are unattended, visual priming is still found for objects presented in the same view, but completely abolished after view changes (see Thoma and Davidoff, 2007, for a brief review). **Hummel (2001) therefore proposed a hybrid model, in which part-based representations are established with attention, whereas view-based representations are automatically activated irrespective of attention.**

Inspired by these previous studies and theoretical considerations, the present functional magnetic resonance imaging (fMRI) study examined the representational format of objects in high-level visual cortex and its dependence on attention. Objects were presented in either an intact or a split configuration (Fig. 1B) and were either attended or unattended. The half-split manipulation, while preserving the constituent object parts, distorted the holistic image in a way that cannot be recovered by the aligning processes of view-based models (Hayward et al., 2010; Thoma et al., 2004). **To prevent verbalization of the attended object as a confounding factor we used a non-semantic attention task, in which participants detected brightness changes on either the object (attended condition) or a contralaterally presented noise stimulus (unattended condition).** We reasoned that only if objects are coded as part-based, non-holistic representations, should activation patterns of split objects be informative about those of intact objects. Moreover, if attention was necessary for part-based representations, this configural invariance of object representations should only be observed for attended, but not for unattended objects.

To this end, we used a novel multivariate approach, in which we trained a support vector machine classifier to discriminate between activation patterns of intact

objects and tested its predictive capacity for activation patterns of split objects and vice versa. The rationale was that successful generalization between activation patterns of intact and split objects is indicative of a non-holistic format of object representations. Our multivariate approach represents a critical advance compared with previous repetition suppression studies, because of mounting evidence that high-level visual areas code objects in a distributed fashion across multiple neuronal populations (Haxby et al., 2001; Rice et al., 2014). Importantly, different configurations of an object might activate identical neuronal populations and the difference between configurations only emerges at the pattern level as a distinct weighting of each population. Multivariate methods are able to pick up on these object- or view-specific multivoxel fingerprints, whereas repetition suppression—as a univariate technique—misses out on such pattern-related information. We focused our analyses on the LOC, given a large body of evidence supporting its pivotal role in object processing (Grill-Spector et al., 1998; Malach et al., 1995) and object recognition (Grill-Spector et al., 2000).

2. Materials and methods

2.1. Participants

Eighteen healthy participants (11 female, mean age ± SEM, 23.4 ± 0.8 years) participated in the experiment for payment after giving written informed consent. The study was conducted according to the declaration of Helsinki, and approved by the local ethics committee.

2.2. Experimental design

The experimental design comprised the factors configuration (intact, split) and attention (attended, unattended) as factors of interest as well as object (camera, watering can, chair) and side of presentation (left, right) as factors of no interest. Within each of 8 experimental runs, an object appeared in 4 trials in each attention conditions (in 2 trials per side of presentation). The order of presentation was randomized across the 48 trials of each run.

2.3. Experimental procedures

A trial (Fig. 1A) started with a blank fixation screen for 3300ms ± 2000ms, after which one half of a central black fixation diamond turned red, indicating the side to which attention should be directed. After a fixed interval (250ms), four repetitions of the stimulus-response phase appeared. Each stimulus-response phase lasted 1500ms and comprised the presentation of the stimulus screen (500ms), a pattern mask (133ms) and the default screen (867ms). An intact or split object appeared on one side of the fixation cross (offset 3.84 degrees of visual angle) and a noise stimulus at the same offset on the other side of the stimulus screen. Intact and split objects as well as the noise stimuli subtended 3.81 by 3.81 degrees of visual angle. A

brightness change occurred 283ms after stimulus onset simultaneously on both the object and the noise stimulus, such that they became independently and randomly either r darker or lighter. Participants were instructed to press a button on the response box when the stimulus on the cued side became darker. The cued stimulus could either be an intact or split image version of an object (*attended condition*) or the noise stimulus (*unattended condition).* Responses were counted as valid within a time window of 1000ms after stimulus offset. **In each repetition of the stimulus-response phase, the same object was shown at the same position. The noise stimulus, while also presented at the same position, was randomly generated for each repetition.**

<Fig. 1 approximately here>

To independently identify object-responsive regions of lateral occipital complex (LOC) in each participant (Malach et al., 1995), we conducted a localizer run with 5 blocks of intact objects, 5 blocks of split objects and 10 blocks of grid-scrambled versions of the objects in randomized order. Blocks lasted for 15.8s during which 20 images were presented for 600ms each, followed by 200ms blank screen. Pairs of identical objects were shown left and right of fixation, equaling the configuration of the main experiment in eccentricity and size. Participants performed a one-back task, in which they had to indicate via button press whenever the same stimulus display appeared twice in a row

2.4. Stimuli

Stimuli were generated with Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) and projected with a Sanyo LCD projector at 60 Hz. The stimulus set consisted of three grayscale objects (camera, watering can, chair) based on realistic three-dimensional models presented

either intact or half-split (Fig. 1B). The objects were selected for representing non-overlapping man-made categories to increase the discriminability of evoked neuronal activation patterns. Split versions were generated by relocating the two halves of an original image to the opposite side of the canvas. The noise stimuli matched the objects in terms of spatial extent and complexity to ensure that there would be no performance difference. They were randomly generated for each trial by sampling a 9 by 9 random binary matrix, scaling the matrix to 216 by 216 pixels, applying a low-pass filter with a cut-off frequency of 0.02/pixel and cropping pixels outside a circle of 216 pixels diameter. This procedure resulted in circular grayscale stimuli with randomly distributed smooth patches. Both the objects and the noise stimuli were scaled to grayscale RGB values between 50 and 205. To generate brightness changes, the underlying RGB histograms were shifted up or down by 50 (the image background remained constant with an RGB value of 200). The pattern masks were generated for each trial by sampling an 18 by 18 random binary matrix and scaling the matrix to 216 by 216 pixels.

2.5. Eyetracking

Eyetracking data were successfully collected in 16 of 18 subjects using an infrared video eyetracking system (iView XTM MRI 50Hz, SensoMotoric Instruments, Teltow, Germany). As a measure of fixation reliability, we computed the percentage of recorded eye gaze positions within a 1.93° visual angle circle around the center of the fixation cross. This radius corresponded to the eccentricity of the inner edges of the two stimulus-containing boxes (see Fig. 1A).

2.6. FMRI data acquisition and preprocessing

FMRI data were acquired on a 3-Tesla Siemens Trio (Erlangen, Germany) scanner using a gradient echo planar imaging (EPI) sequence and a 12-channel head-coil. We recorded 8 experimental runs of 214 whole-brain volumes each (TR = 2s, echo time (TE) 25 ms, flip angle 78°, 33 slices, 3mm isotropic resolution, interslice gap 0.75mm). The LOC localizer comprised 242 volumes. In addition, a high-resolution T1-weighted image was acquired (TR = 1.9s, echo time (TE) 2.51 ms, flip angle 9°, 192 slices, resolution 1mm isotropic). Preprocessing was performed using SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London) and included realignment and smoothing with an 8mm Gaussian kernel. All main analyses were performed in native subject space.

2.7. Region of interest procedures

Our main region of interest (ROI) was LOC. To anatomically constrain LOC, which stretches from lateral occipital cortex to posterior fusiform gyrus (Grill-Spector et al., 1999), we generated a bilateral composite mask of the inferior occipital cortex, middle occipital cortex and the posterior half of the fusiform gyrus (derived from the AAL Atlas, Tzourio-Mazoyer et al., 2002). The LOC ROI was defined as the intersection of the anatomical mask and the functional localizer based on the group-level T-contrast *intact + split > scrambled* at a significance level of $p<0.05$ (family-wise error (FWE) corrected at the whole-brain level). Additionally we created separate ROIs for two subregions of LOC, lateral occipital cortex (LO; corresponding to the inferior and middle occipital anatomical masks) and posterior fusiform gyrus (pFus; posterior fusiform gyrus mask), based on previous reports regarding a possible functional dissociation between the two (Cichy et al., 2013; Grill-Spector et al.,

2001). The V1 ROI was defined as the intersection of Brodmann Area 17 (derived from the SPM Anatomy toolbox; Eickhoff et al., 2005) and the functional localizer based on the group-level T-contrast *intact + split + scrambled > implicit baseline* at a significance level of $p_{FWE}<0.05$. The V1 ROI was based on a mask for Brodmann Area 17 derived from the SPM Anatomy toolbox (Eickhoff et al., 2005). All ROIs were reverse-normalized to native subject space.

2.8. FMRI data analysis

2.8.1.First-level general linear models (GLMs)

For each participant we estimated a GLM with separate experimental regressors for the factors configuration (split, intact), attention (attended, unattended), object (camera, watering can, chair) and side of presentation (left, right). Onsets of the experimental regressors were set to the beginning of the stimulus-response phase, and they were modeled as stick functions and convolved with a canonical hemodynamic response function. Further, six motion parameters from the realignment preprocessing step were included as regressors-of-no-interest.

The GLM for the functional localizer comprised regressors for intact objects, split objects and scrambled objects and six motion parameters. The experimental regressors were modeled as boxcar functions with durations equal to the block lengths (15.8s) and convolved with a canonical hemodynamic response function.

2.8.2.Multivariate analyses

The estimated beta images of the GLM provided the basis for support vector machine (SVM)

classification. SVM classification was performed using *The Decoding Toolbox* (Görgen et al., 2012) with a linear C-SVM and a fixed cost parameter (c=1).

We first performed a searchlight analysis (Kriegeskorte et al., 2006), in which a sphere with a radius of 4 voxels was centered at each voxel of the brain and decoding was based on the voxels within each sphere. A leave-one-run-out cross-validation procedure was used, such that in each fold the classifier was trained on the beta maps of seven runs and tested on the left out eighth run. The resulting decoding accuracies were assigned to the center voxel. We performed decoding separately between the three pairs of objects (camera-can, camera-chair, can-chair) in each of the four experimental conditions (intact attended, intact unattended, split attended, split unattended) and both sides of presentation (left, right). After averaging across object pairs and sides, we obtained information maps for each subject and experimental condition, which were subsequently normalized to a common template for group-level statistical inference.

The main analyses were based on ROI decoding (Haynes and Rees, 2005; Kamitani and Tong, 2005), in which the voxels of a given ROI in native space were used for classification. ROI decoding followed the same cross-validation procedure as detailed for the searchlight analysis**. In addition, we used a nested feature selection procedure in order to select the most stimulus-responsive voxels. Thus, for each of the seven runs within a fold of the cross-validation procedure, voxels were ranked according to the magnitude (beta value) of the stimulus-related responses in the respective six other runs. Stimulus-related responses were derived from the T-contrast *all conditions > implicit baseline*.**

We refer to decoding analyses in which training and testing was performed within the

same configuration (e.g. training on intact objects, testing on intact objects) as *within-configuration* decoding. In the *cross-configuration* analysis we trained the classifier to discriminate between intact object categories and tested on split object categories and vice versa, and then averaged across train-test directions. In the *cross-attention* analysis the classifier was trained to discriminate between attended object categories and tested on unattended object categories and vice versa. We performed *cross-attention* classification both under the *within-* and the *cross-configuration* decoding scheme as described above.

For statistical inference we performed two-sided t-tests and repeated-measures ANOVAs. Two-tailed t-tests for decoding accuracies were tested against the null hypothesis of a chance level decoding performance of 50%.

3. Results

3.1. Behavioral results and fixation control

Participants detected and reported brightness changes of the objects and the noise stimuli highly accurately (performance > 98%), indicating that they focused their attention on the correct stimulus in each condition. On average, 98.3±0.8% of recorded eye gaze positions were within the fixation area, demonstrating that the participants maintained fixation throughout the experiment.

3.2. Decoding of objects

Initially, we performed a searchlight analysis to identify brain areas that processed information about object categories (Fig. 2). We found above-chance classification for split and intact object in both the attended and the unattended condition in areas overlapping with the three a priori defined ROIs (LO, pFus,V1; peaks in the three ROIs were significant for all conditions at $p<0.01$, FWE-corrected for small volumes). We did not observe significant above-chance decoding beyond the predefined ROIs. We therefore performed all subsequent analyses in those ROIs. Further, since we did not find any differences between pFus and LO in any of the following analyses, we present the results for a composite mask of pFus and LO (LOC ROI).

<Fig. 2 approximately here>

ROI decoding accuracies were significantly above chance in all conditions in LOC and V1 (Table 1). Repeated-measures ANOVA (rmANOVA) with the factors attention and configuration showed a main effect of attention in LOC, ($F(1,17)=39.4$, $p<0.001$), but not in

V1 ($F(1,17)=2.0$, $p=0.17$). The strong main effect of attention demonstrates the effectiveness of the attentional manipulation. No other main effects or interactions were significant in either ROI (all $p>0.1$).

<Table 1 approximately here>

3.3. Generalization between intact and split object representations

The critical test for distinguishing formats of object representations **in LOC** was whether the classifier generalized between intact and split objects. For this we conducted a *cross-configuration* analysis, training the classifier on intact objects and testing on split objects, and vice versa. In LOC, *cross-configuration* decoding of objects was significant in both the attended (67.3% accuracy, $t(17)=8.9$, $p<0.001$) and unattended condition (55.5% accuracy, $t(17)=3.7$, $p=0.002$; see Fig. 3A). This generalization clearly suggests a non-holistic format of object representations **in LOC**.

To assess how *cross-configuration* decoding compared with *within-configuration* decoding, we conducted an rmANOVA with factors decoding scheme and attention. There was a main effect of attention ($F(1,17) = 43.1$, $p<0.001$), but neither a main effect of decoding scheme ($F(1,17) = 0.01$, $p=0.91$) nor an attention-by-decoding scheme interaction ($F(1,17)=0.1$, $p=0.70$). Thus the classifier could equally well predict intact and split objects, irrespective of whether it was trained on intact or split objects (configural invariance).

In V1, by contrast, we found a main effect of decoding scheme ($F(1,17)=71.8$, $p<0.001$), such that *within-configuration* decoding was superior to *cross-configuration* decoding. There was neither a main effect of attention ($F(1,17) = 0.33$, $p=0.57$) nor an

attention-by-decoding scheme interaction ($F(1,17) = 3.6$, $p=0.08$). An rmANOVA with the additional factor region revealed a region-by-decoding scheme interaction ($F(1, 17) = 32.0$, $p<0.001$), showing that the observed configural invariance was present in LOC, but not V1 (Fig 3A). The post-hoc probability to detect a main effect of decoding scheme in LOC of the same effect size as in V1 was 0.978.

<Fig. 3 approximately here>

To test whether the generalization between intact and split objects in LOC would also hold if they were presented in different hemifields, we repeated the above analyses in a cross-hemifield decoding scheme. The classifier was trained on stimuli in one hemifield and tested on stimuli in the other hemifield, both under a within- and cross-configuration scheme. As shown in Supplementary Fig. S1A <Insert Supplementary Figure S1 here>, cross-configuration decoding was significant for attended objects (54.8% accuracy, $t(17)=5.1$, $p<0.001$) and at the same level as within-configuration decoding (54.6% accuracy, $t(17)=5.6$, $p<0.001$). The analysis **demonstrates** that the finding of complete cross-configuration generalization persists for high-level neuronal populations with receptive fields encompassing an area of at least 5.7 degree visual angle left and right of fixation (11.4 degree in total). Due to insufficient power, no statement can be made about unattended objects (Supplementary Figure S1B).

3.4. Generalization between attended and unattended object representations

Our finding that representations of both attended and unattended objects in LOC were insensitive to the split procedure does not preclude the possibility that LOC relies on different neural representations for attended and unattended objects. We therefore trained a classifier in

a *cross-attention/within-configuration* analysis on attended objects and tested on unattended objects (and vice versa). The classifier was able to cross-classify activation patterns of attended and unattended objects (57.1% accuracy, $t(17)=5.9$, $p<0.001$) in LOC (Fig. 3B), strongly suggesting that attended and unattended objects share a common representational basis. Furthermore, *cross-attention* decoding was successful even under a *cross-configuration* decoding scheme (57.1% accuracy, $t(17)=6.7$, $p<0.001$; Fig. 3B), demonstrating that the information shared between attended and unattended object representations in LOC is coded non-retinotopically.

Cross-attention/within-configuration decoding was also successful in V1 (57.2% accuracy, $t(17)=13.2$, $p<0.001$; Fig. 3B), while *cross-attention/cross-configuration* decoding was not (48.8% accuracy, $t(17)=-1.0$, $p=0.36$). This difference was significant ($t(17)=7.2$, $p<0.001$). Across regions (V1, LOC) there was a region-by-decoding scheme interaction ($F(1,17) = 31.9$, $p<0.001$), consistent with the presence of configural invariance at the level of LOC, but not at the level of V1, found in the *within-attention* analysis.

4. Discussion

We investigated the representational format of objects **in LOC** and its relation to attention. We found that activation patterns of intact and split objects shared information that allowed mutual prediction of the presented object (*cross-configuration* decoding). Remarkably, *cross-configuration* decoding was at the same level as *within-configuration* classification, that is, intact objects representations were predicted by activation patterns of split objects just as well as by activation patterns of intact objects and vice versa. **Crucially, cross-configuration decoding did not depend on attention.** This pattern of results strongly suggests that the representational code of objects in LOC is based on a non-holistic format **irrespective of attention**.

Previous studies showed that BOLD activity in the LOC is barely affected when objects are—as in our study—coarsely scrambled (half-split and two-fold splits: Lerner et al., 2001; 8-fold: Grill-Spector et al., 1998), whereas finer scrambling leads to a strong reduction (Grill-Spector et al., 1998; Lerner et al., 2001). However, whether the activation elicited by coarsely scrambled objects still contains meaningful object-related information remained unclear. We extended those findings by showing that the information coded in activation patterns of intact objects is preserved in the patterns of split objects. Our results in combination with the above reports therefore support a model in which LOC codes objects as part-based representations.

A part-based coding scheme is in line with behavioral priming studies reporting mirror (Biederman and Cooper, 1991), rotational (Biederman and Gerhardstein, 1993) and configural (Thoma et al., 2004) invariance of object representations. Part-based models,

which posit independent encoding of object parts, correctly predict priming effects for the range of manipulations above, because the constituent object parts are preserved between prime and probe displays. A number of previous studies have probed the viewpoint dependence of object representations at the neural level by examining repetition suppression (RS). Some of those studies found evidence for viewpoint-invariant representations in high-level visual cortex (Eger et al., 2004; James et al., 2002; Kourtzi et al., 2003), others found tolerance only in the left hemisphere (Vuilleumier et al., 2002) or not at all (Grill-Spector et al., 1999). While the present study cannot reconcile those reports, it introduces a new, multivariate perspective to the longstanding question of view dependence. Our cross-classification approach is sensitive to object-related information coded at the level of multivoxel activation patterns, which could not be assessed by previous imaging studies using univariate fMRI data analysis. At the level of multivoxel activation patterns we found a striking invariance of object representations with respect to the relative dislocation of object parts in LOC, indicative of a part-based code with all its theoretical advantages regarding robust and flexible coding of objects under various viewing conditions (Hummel and Biederman, 1992).

It should be noted that our finding of cross-configuration generalization is not informative about the specific nature or complexity of object parts. For instance, since our split manipulation only distorted the overall holistic image but not individual parts, our results are open to the possibility that the part representations themselves are view-based. However, our finding of relative position invariance of object parts entails important constraints on models of object recognition. The "chorus of fragments" model by Edelman and Intrator (2000), for instance, poses "what+where" units coding the conjunction of part

(fragment) information and retinotopic position, which is at odds with relative position invariance of object parts. Our results seem to fit however a "bag of features" (Hayworth et al., 2011) model, as for instance proposed by Ullman and colleagues (Epshtein and Ullman, 2007; Ullman, 2007).

Our second key finding is that the format of object representations in LOC was independent of attention. We could cross-classify between intact and split objects, whether they were attended or unattended. This finding was further corroborated by the fact that we were able to predict the attended objects based on the activation patterns of unattended objects (and vice versa). Thus, not only appears the LOC to adhere to a part-based format irrespective of attention, but the underlying neural representations additionally seem to be shared between attended and unattended objects. Importantly, we found this *cross-attention* generalization also under a *cross-configuration* decoding scheme, demonstrating that the activation patterns of attended and unattended objects indeed shared non-retinotopic, high-level information. **Taken together, our results do not provide evidence for a critical role of attention for part-based representations, as implicated by other empirical findings (Stankiewicz et al., 1998; Thoma and Henson, 2011; Thoma et al., 2004) and theoretical accounts (Hummel, 2001).** The main difference between attended and unattended object representations in our study was of quantitative nature—superior decoding accuracy for attended objects, likely related to neural gain (see Pratte et al., 2013)—but not qualitative in terms of the underlying representational format**.** Although, **given its focus on neural effects, our study is not in the position to challenge the findings from these previous behavioral priming studies,** it has a number of important advantages. First, we ensured the effectiveness of our attentional manipulation both at the behavioral—by means of eyetracking—and at the

neural level, based on superior decoding accuracies for attended relative to unattended objects. Second, we directly probed the representational format **in LOC** and its dependence on attention, whereas **behavioral** priming studies rely on indirect inference from reaction times. And third, we used a task (brightness discrimination) that attenuates the semantic aspect of object recognition, which is arguably more pronounced in the object naming tasks employed by many priming studies. Semantic top-down feedback from higher areas is a potential confounding factor in these studies, since feedback might be responsible for view-invariant priming, but might itself be dependent on attention. **Future studies could investigate whether our finding of configurational invariance for both attended and unattended object representations is conditional on using a non-semantic object perception task, or whether it holds for tasks requiring object identification.**

Our results also differ from a previous neuroimaging study that found evidence for a part-based format in LOC only for attended, but not for unattended, objects (Thoma and Henson, 2011). The paradigm and the analyses of Thoma and Henson deviate in a number of important aspects from those of the present study. The presentation times in Thoma and Henson were considerably shorter than in our study (135 milliseconds, compared to 2 seconds in our study), which opens the possibility that the instantiation of part-based representations requires longer presentations times in the absence of attention. This assertion would be consistent the notion that attention boosts neuronal processing of stimuli by increasing the signal-to-noise ratio of neuronal responses (Bisley, 2011). Another noteworthy difference is the fact that in our study objects were repeated multiple times throughout the experiment, whereas each object in Thoma and Henson appeared in exactly one trial. Under the assumption of a view-based

object format, the multiple repetitions in our study could have therefore led to the instantiation of view-based representations of (previously unfamiliar) split objects. However, the fact that we found perfect generalization between intact and split objects, despite the large number of repetitions, argues against a build-up (or prior existence) of view-based representations for split objects. On the methodological side, Thoma and Henson assessed the effects of RS, whereas we employed a cross-classification approach that utilized the full pattern information. Given that unattended objects evoke a weaker BOLD response (Murray and Wojciulik, 2004; O'Craven et al., 1999; Serences et al., 2004), it seems possible that RS was not sensitive enough to detect representational commonalities between intact and split unattended objects. Additionally, the analysis of repetition suppression effects by Thoma and Henson misses out on information coded at the pattern level, which by itself could explain the observed discrepancies for unattended objects.

Finally, an important consideration is that the support vector machine approach in our study is largely intransparent with respect to the particular stimulus features underlying successful classification. A possible concern could be that between-object decoding might have entirely been based on low-level visual features. For instance, the surface of an object might have a certain characteristic texture, and further, the same kind of texture might even be present at a similar retinotopic location after the splitting procedure—hence explaining the observed cross-configuration generalization. However, for several reasons we consider a pure low-level account of our results unlikely. First, if retinotopic low-level features were an important source of discriminative information between objects, cross-configuration decoding should have worked in V1 as well, which it did not. Second, the intact and split

versions of our object images had very little low-level commonalities in retinotopic coordinates as an image analysis with a biologically plausible model of visual cortex confirmed (Supplementary Fig. S2) <Insert Supplementary Figure S2 here>. **Third, the voxels entering the multivariate analysis were precisely selected for preferring complex features over low-level features.** And fourth, the results of the cross-hemifield analysis show that the cross-configuration generalization holds also for high-level representations with near-complete location invariance (Supplementary Fig. S1). Therefore, while acknowledging the possibility that the classifier could have only picked up on low-level information, we consider such an account highly unlikely for the reasons outlined.

In summary, our study provides novel evidence indicating that neural representations of both attended and unattended objects in LOC rely on a non-holistic rather than view-based format. Moreover, our data strongly suggest that attended and unattended objects rely on a common representational format.

Acknowledgements

This research was supported by the German Research Foundation (DFG) through the Research Training Group GRK1589/1 (to M.G. and P.S.), and Grants STE1430/6-1 (to P.S.), and RI1847/1-1 and SFB779TPA10N (to A.R.-K). R.C. was founded by a Feodory Lynen Grant of the Alexander von Humboldt Foundation. We thank Guy Middleton for assistance with rendering the object images from 3D models.

References

Andresen, D.R., Vinberg, J., Grill-Spector, K., 2009. The representation of object viewpoint in human visual cortex. Neuroimage 45, 522–536.

Biederman, I., Cooper, E.E., 1991. Evidence for complete translational and reflectional invariance in visual object priming. Perception 20, 585–593.

Biederman, I., Gerhardstein, P.C., 1993. Recognizing Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint In variance. J. Exp. Psychol. Hum. Percept. Perform. 19, 1162–1182.

Bisley, J.W., 2011. The neural basis of visual attention. J. Physiol. 589, 49–57.

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436.

Bülthoff, H.H., Edelman, S., 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc. Natl. Acad. Sci. U. S. A. 89, 60–64.

Cichy, R.M., Sterzer, P., Heinzle, J., Elliott, L.T., Ramirez, F., Haynes, J.-D., 2013. Probing principles of large-scale object representation: category preference and location encoding. Hum. Brain Mapp. 34, 1636–1651.

Edelman, S., Bülthoff, H.H., 1992. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Res. 32, 2385–2400.

Edelman, S., Intrator, N., 2000. (Coarse coding of shape fragments) Representation of structure. Spat. Vis. 13, 255–264.

Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R.J., Rees, G., 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. J. Cogn. Neurosci. 20, 356–370.

Eger, E., Henson, R.N.A., Driver, J., Dolan, R.J., 2004. BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. J. Neurophysiol. 92, 1241–1247.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.

Epshtein, B., Ullman, S., 2007. Semantic Hierarchies for Recognizing Objects and Parts, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR ’07. pp. 1–8.

Ewbank, M.P., Schluppeck, D., Andrews, T.J., 2005. fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. Neuroimage 28, 268–279.

Gauthier, I., Hayward, W.G., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 2002. BOLD Activity during Mental Rotation and Viewpoint-Dependent Object Recognition. Neuron 34, 161–171.

Görgen, K., Hebart, M.N., Haynes, J.-D., 2012. The Decoding Toolbox (TDT): a new fMRI analysis package for SPM and MATLAB. Conf. Hum. Brain Mapp. Organ. Beijing, China.

Grill-Spector, K., Henson, R., Martin, A., Grillspector, K., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10, 14–23.

Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. Vision Res. 41, 1409–1422.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24, 187–203.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., Malach, R., 1998. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum. Brain Mapp. 6, 316–328.

Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. Nat. Neurosci. 3, 837–43.

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, 2425–2430.

Haynes, J.-D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691.

Hayward, W.G., Zhou, G., Man, W.-F., Harris, I.M., 2010. Repetition blindness for rotated objects. J. Exp. Psychol. Hum. Percept. Perform. 36, 57–73.

Hayworth, K.J., Biederman, I., 2006. Neural evidence for intermediate representations in object recognition. Vision Res. 46, 4024–4031.

Hayworth, K.J., Lescroart, M.D., Biederman, I., 2011. Neural encoding of relative position. J. Exp. Psychol. Hum. Percept. Perform. 37, 1032–1050.

Hummel, J.E., 2001. Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. Vis. cogn. 8, 489–517.

Hummel, J.E., Biederman, I., 1992. Dynamic binding in a neural network for shape recognition. Psychol. Rev. 99, 480–517.

James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., Goodale, M.A., 2002. Differential Effects of Viewpoint on Object-Driven Activation in Dorsal and Ventral Streams. Neuron 35, 793–801.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

Kourtzi, Z., Erb, M., Grodd, W., Bülthoff, H.H., 2003. Representation of the perceived 3-D object shape in the human lateral occipital complex. Cereb. cortex 13, 911–920.

Kourtzi, Z., Kanwisher, N., 2000. Cortical regions involved in perceiving object shape. J. Neurosci. 20, 3310–3318.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. 103, 3863–3868.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., Malach, R., 2001. A hierarchical axis of object processing stages in the human visual cortex. Cereb. Cortex 11, 287–97.

Logothetis, N.K., Pauls, J., Bülthoff, H.H., Poggio, T., 1994. View-dependent object recognition by monkeys. Curr. Biol. 4, 401–414.

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. 92, 8135–8139.

Marr, D., Nishihara, H.K., 1978. Representation and recognition of the spatial organization of three-dimensional shapes. Proc. R. Soc. London. Ser. B, Biol. Sci. 200, 269–294.

Murray, J.E., 1999. Orientation-specific effects in picture matching and naming. Mem. Cognit. 27, 878–889.

Murray, S.O., Wojciulik, E., 2004. Attention increases neural selectivity in the human lateral occipital complex. Nat. Neurosci. 7, 70–74.

O'Craven, K.M., Downing, P.E., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. Nature 401, 584–587.

Olshausen, B.A., Anderson, C.H., Essen, D.C. Van, 1993. A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information. J. Neurosci. 13, 4700–4719.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442.

Perrett, D.I., Oram, M.W., Ashbridge, E., 1998. Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. Cognition 67, 111–145.

Poggio, T., Edelman, S., 1990. A Network that Learns to Recognize 3D Objects. Nature 343, 263–266.

Pratte, M.S., Ling, S., Swisher, J.D., Tong, F., 2013. How attention extracts objects from noise. J. Neurophysiol. 110, 1346–1356.

Rice, G.E., Watson, D.M., Hartley, T., Andrews, T.J., 2014. Low-Level Image Properties of Visual Objects Predict Patterns of Neural Response across Category-Selective Regions of the Ventral Visual Pathway. J. Neurosci. 34, 8837–8844.

Serences, J.T., Schwarzbach, J., Courtney, S.M., Golay, X., Yantis, S., 2004. Control of object-based attention in human cortex. Cereb. cortex 14, 1346–1357.

Stankiewicz, B.J., Hummel, J.E., Cooper, E.E., 1998. The role of attention in priming for left-right reflections of object images: evidence for a dual representation of object shape. J. Exp. Psychol. Hum. Percept. Perform. 24, 732–744.

Tarr, M.J., 1995. Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psychon. Bull. Rev. 2, 55–82.

Tarr, M.J., Gauthier, I., 1998. Do viewpoint-dependent mechanisms generalize across members of a class? Cognition 67, 73–110.

Tarr, M.J., Pinker, S., 1989. Mental rotation and orientation-dependence of shape recognition. Cogn. Psychol. 21, 233–282.

Thoma, V., Davidoff, J., 2007. Object recognition: attention and dual routes, in: Osaka, N., Rentschler, I., Biederman, I. (Eds.), Object Recognition, Attention, and Action. Springer, Tokyo, pp. 141–158.

Thoma, V., Henson, R.N., 2011. Object representations in ventral and dorsal visual streams: fMRI repetition effects depend on attention and part-whole configuration. Neuroimage 57, 513–525.

Thoma, V., Hummel, J.E., Davidoff, J., 2004. Evidence for holistic representations of ignored images and analytic representations of attended images. J. Exp. Psychol. Hum. Percept. Perform. 30, 257–267.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.

Ullman, S., 1989. Aligning pictorial descriptions: An approach to object recognition. Cognition 32, 193–254.

Ullman, S., 1998. Three-dimensional object recognition based on the combination of views. Cognition 67, 21–44.

Ullman, S., 2007. Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn. Sci. 11, 58–64.

Vuilleumier, P., Henson, R.N., Driver, J., Dolan, R.J., 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat. Neurosci. 5, 491–9.

Captions

Fig. 1. Experimental procedures and stimuli. A. In each trial a cue indicated the side to which attention should be directed. Subsequently, four repetitions of the stimulus-response phase appeared, during each of which participants had to detect a decrease in brightness of either the object (attended condition) or the noise stimulus (unattended condition). B. The stimulus set consisted of three objects in an intact and half-split configuration.

Fig. 2. Searchlight analysis results for intact and split objects based on a *within-configuration*/between-object decoding procedure. Whole-brain information maps are represented as T-maps indicating the statistical significance of voxel-wise decoding accuracies against the chance-level decoding accuracy of 50%. The T-maps are thresholded at $p<0.005$, uncorrected, for illustration. A. Attended objects. B. Unattended objects.

Fig. 3. *Within-* and *cross-configuration* decoding in LOC and V1. A. *Within-attention* decoding scheme. B. *Cross-attention* decoding scheme. Error bars represent SEM. P-values are based on two-tailed paired t-tests. Stars represent the significance of decoding accuracies based on two-tailed t-tests against the chance-level decoding accuracy of 50%: ** $p<0.01$ *** $p<0.001$.

Table 1. Basic ROI decoding results in LOC and V1 for intact and split objects and for both the attended and unattended condition.

Tables

Table 1

		accuracy	t(17)	p
Intact attended	**LOC**	66.9%	9.8	.00000002
	V1	63.3%	8.9	.00000008
Split attended	**LOC**	67.9%	6.8	.000003
	V1	62.4%	8.0	.0000004
Intact unattended	**LOC**	54.1%	2.2	.043
	V1	59.3%	5.7	.00003
Split unattended	**LOC**	55.9%	3.6	0.002
	V1	61.6%	5.2	0.00007

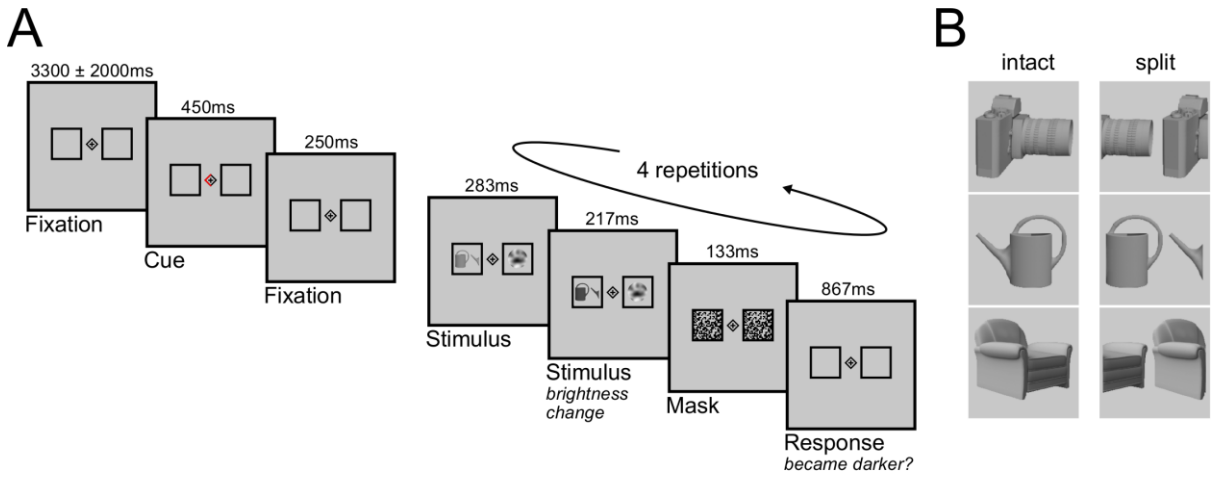

A
3300 ± 2000ms
Fixation
450ms
Cue
250ms
Fixation
4 repetitions
283ms
Stimulus
217ms
Stimulus
brightness
change
133ms
Mask
867ms
Response
became darker?
B
intact
split

Figure 1

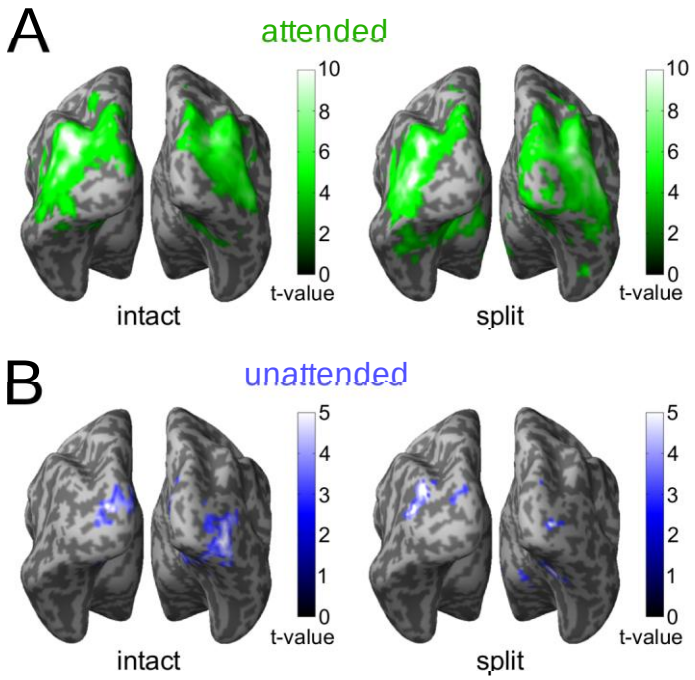

A
attended
10
8
6
4
2
0
t-value
intact
split
B
unattended
5
4
3
2
1
0
t-value
intact
split

Figure 2

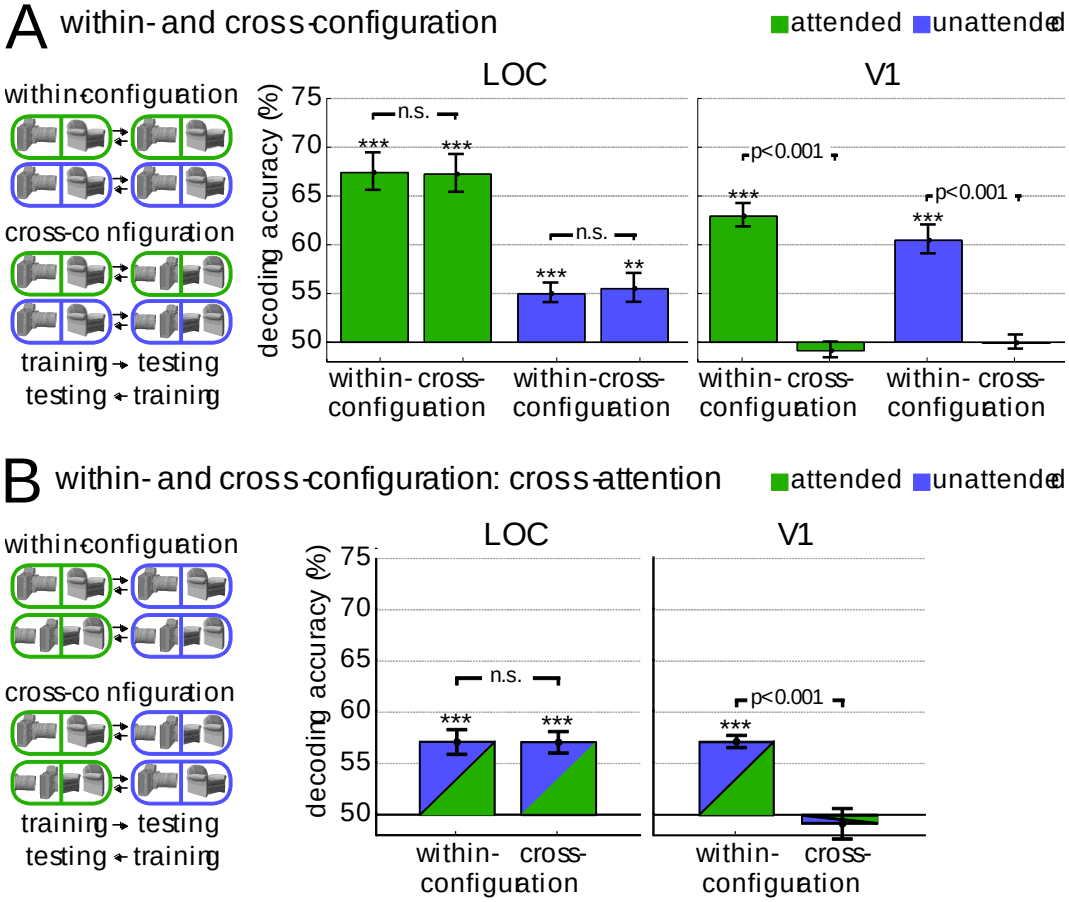
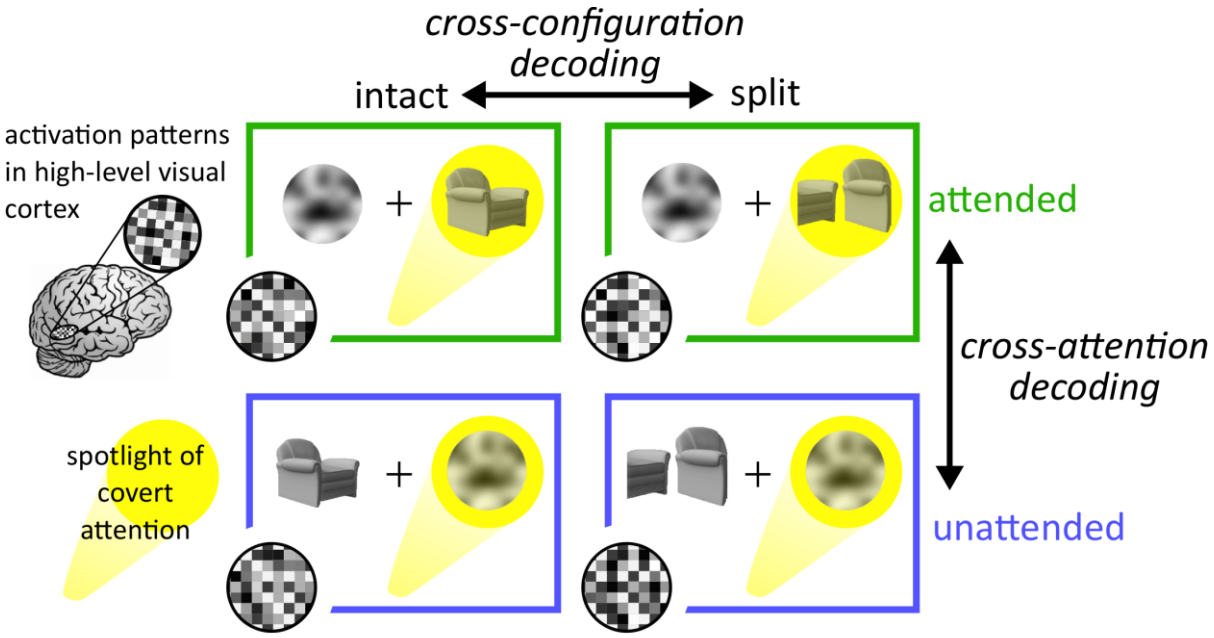

A within- and cross-configuration
attended
unattended
within-configuration
cross-configuration
training → testing
testing → training
LOC
V1
decoding accuracy (%)
75
70
65
60
55
50
n.s.
***
***
n.s.
***
**
p<0.001
***
p<0.001
***
within-configuration
cross-configuration
within-configuration
cross-configuration
within-configuration
cross-configuration
within-configuration
cross-configuration
B within- and cross-configuration: cross-attention
attended
unattended
within-configuration
cross-configuration
training → testing
testing → training
LOC
V1
decoding accuracy (%)
75
70
65
60
55
50
n.s.
***
***
p<0.001
***
within-configuration
cross-configuration
within-configuration
cross-configuration

Figure 3



Graphical abstract

Highlights:

- Non-holistic coding of objects in LOC
- Relative position invariance of object parts in LOC
- No evidence for a role of attention in establishing a non-holistic code in LOC
- Common neural basis of attended and unattended objects in LOC