Disentangling visual imagery and perception of real-world objects

Sue-Hyun Lee, **Dwight J. Kravitz**, and **Chris I. Baker**
Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892

Abstract

During mental imagery, visual representations can be evoked in the absence of "bottom-up" sensory input. Prior studies have reported similar neural substrates for imagery and perception, but studies of brain-damaged patients have revealed a double dissociation with some patients showing preserved imagery in spite of impaired perception and others vice versa. Here, we used fMRI and multi-voxel pattern analysis to investigate the specificity, distribution, and similarity of information for individual seen and imagined objects to try and resolve this apparent contradiction. In an event-related design, participants either viewed or imagined individual named object images on which they had been trained prior to the scan. We found that the identity of both seen and imagined objects could be decoded from the pattern of activity throughout the ventral visual processing stream. Further, there was enough correspondence between imagery and perception to allow discrimination of individual imagined objects based on the response during perception. However, the distribution of object information across visual areas was strikingly different during imagery and perception. While there was an obvious posterior-anterior gradient along the ventral visual stream for seen objects, there was an opposite gradient for imagined objects. Moreover, the structure of representations (i.e. the pattern of similarity between responses to all objects) was more similar during imagery than perception in all regions along the visual stream. These results suggest that while imagery and perception have similar neural substrates, they involve different network dynamics, resolving the tension between previous imaging and neuropsychological studies.

Keywords

Object Recognition; fMRI; Human; Mental Imagery; Vision; Top-down

Introduction

Our everyday visual perception reflects the interaction of externally driven "bottom-up" sensory information and internally generated "top-down" signals, which guide interpretation of sensory input (Hsieh et al., 2010; Kastner et al., 1998). However, even in the absence of bottom-up signals it is still possible to generate internal visual representations using top-down signals only, commonly referred to as mental imagery. Here we investigate the extent

to which imagery shares the same neural substrate and mechanisms with perception in the ventral visual pathway (Farah, 1999; Kosslyn et al., 2001; Pylyshyn, 2002).

Prior studies using psychophysics (Ishai and Sagi, 1995; Pearson et al., 2008; Winawer et al., 2010), functional brain imaging (Ganis et al., 2004; Kosslyn et al., 1997) and intracranial recordings (Kreiman et al., 2000) have suggested similar mechanisms for imagery and perception. For example, the global pattern of brain activation in both imagery and perception is remarkably similar (Ganis et al., 2004; Kosslyn et al., 1997). Further, imagery can elicit category specific responses in high-level visual areas (Ishai et al., 2000; O'Craven and Kanwisher, 2000; Reddy et al., 2010) and retinotopically specific activity in primary visual cortex (Klein et al., 2004; Slotnick et al., 2005; Thirion et al., 2006) Finally, a recent study reported that the identity of one of two possible imagined stimuli ('X' or 'O') could be decoded from the pattern of response elicited by seeing those same stimuli in high-level object-selective cortex, suggesting overlap in the representations evoked during imagery and perception (Stokes et al., 2011; Stokes et al., 2009).

Yet, despite these similarities, imagery and perception are clearly distinct. Seeing and imagining are very different phenomenologically, and studies of brain-damaged individuals suggest that imagery and perception can be dissociated (Bartolomeo, 2002, 2008; Behrmann, 2000). For example, the object agnosic patient CK who has damage in the ventral visual pathway (Behrmann et al., 1994; Behrmann et al., 1992) is unable to recognize objects but can reproduce detailed drawings from memory and has preserved visual imagery. Conversely, deficits in visual imagery have been reported in the absence of agnosia, low-level perceptual deficits or disruption of imagery in other modalities (Farah et al., 1988; Moro et al., 2008). Thus, imagery and perception seem to share a neural substrate but are nevertheless dissociable. Therefore the critical question is what differentiates the utilization of the tissue by the two processes along the ventral visual pathway.

Here, we investigated the *specificity* (discrimination of objects), *distribution* (comparison of visual areas from V1 to high-level visual cortex), and *differences* between visual representations during perception and imagery of 10 individual real-world object images. While most previous studies have used stimuli (e.g. Gabor filters) or comparisons (e.g. category) tailored for specific regions of the ventral visual pathway, we ensured that perceptual decoding was possible across the entire pathway, allowing a systematic comparison of imagery and perception both within and across visual areas. We find that while there are similarities in the representations of seen and imagined objects, allowing decoding during imagery throughout the ventral pathway, there are critical differences. In particular, the distribution of information from V1 to high-level visual cortex showed opposite gradients. Further, the structure of representations (pattern of correlations between object images) across visual areas was more similar during imagery than during perception. Thus, while imagery and perception share at least some of the same neural substrate, they engage distinct neural mechanisms and involve different network dynamics along the ventral visual pathway, providing a potential resolution to the contradiction between prior imaging and neuropsychological studies.

Material and methods

Participants

Eleven neurologically intact, right-handed participants (5 males, 6 females, age 25 ± 1 years) took part in this study (2 additional participants were excluded due to a failure to localize early visual cortical regions). All participants provided written informed consent for the procedure in accordance with protocols approved by the NIH institutional review board.

Stimuli

We used ten images of common objects: bag, car, chair, clock, jet, lamp, necklace, pen, umbrella and violin (Fig. 1A). These objects were chosen to differ widely in their orientation, shape, and color, so that during perception there would be highly distinct patterns of activity in retinotopic visual areas in addition to object-selective areas. Enabling perceptual decoding in both retinotopic and object-selective cortex allows us to compare directly these areas in perception and imagery. However, any decoding in object-selective cortex may still reflect the retinotopic differences in the stimuli rather than high-level form representations.

Pre-scan training

Prior to scanning, participants were familiarized with the ten object images and practiced generating vivid mental images. This training was comprised of three separate phases. In the first phase, participants practiced general mental imagery and scored the subjective vividness of their mental images. Specifically, participants rated the vividness of mental images about specific scenes or situations instructed by "The Vividness of Visual Imagery Questionnaire" (VVIQ)(Cui et al., 2007; Marks, 1973). A rating of 1 indicated the most perfect vivid image, as good as seeing the image, and 5 meant no image (only "know"). In the second phase, participants were familiarized with the set of 10 object images (Fig. 1A). Each object image was presented for 4 s, followed by three questions to encourage participants to think about the visual details (color, shape, pattern, etc.) of the objects. For example: "What kind of pattern is on the umbrella?", "What color is the handle of the umbrella?", and "Which side is the handle facing?". After the questions, there were two further presentations of each of the object images. The final phase of the training session was a practice of the experimental trials (blocked by imagery or perception) and consisted of 3 repetitions of a perception block and an imagery block. During the perception block, participants were presented sequentially with each of the 10 object images combined with auditory presentations of the object names at a rate of 1 image every 3 s in random order. During the imagery block, participants were asked to imagine each of the 10 objects when they heard the names of the object at a rate of 1 object imagery trial every 3 s in random order.

fMRI Experiment

The main task consisted of six runs, each lasting 480 s. Each run involved two trial types, perception and imagery, presented in a fully interleaved event-related fashion. Participants were asked to maintain fixation on a central cross throughout each run. Mid-level gray background was used for all presentations. On each trial, the white fixation cross first changed color indicating either a perception trial (blue) or an imagery trial (red). After 500 ms, participants heard a 500 ms long recording of the spoken name of the object to be presented or imagined on that trial. In the perception condition, the auditory cue was immediately followed by a 2 s presentation of the object corresponding to the cue from among the 10 objects seen during training. In the imagery condition, participants were instructed to imagine for 2 s the specific image of the object given by the auditory cue when they saw the red fixation cross. Each trial lasted 3s with a variable inter-trial interval of 3 – 13 s. During each run, every object occurred in 3 perception and 3 imagery trials, for a total of 60 trials per run. The order of the conditions and objects were randomized and counterbalanced across runs (Fig. 1B).

Localizer Design

Each participant completed two localizer scans. The first localizer was used to identify object-selective regions of cortex. Participants viewed alternating 16 s blocks of grayscale

object images and retinotopically matched scrambled images (Kourtzi and Kanwisher, 2000; Kravitz et al., 2010). The resulting object-selective lateral occipital complex (LOC) was divided into a posterior subdivision (lateral occipital; LO) and an anterior subdivision (posterior fusiform, pFs) because it has been suggested that the lateral occipital complex may be spatially and functionally segregated along an anterior-posterior axis (Grill-Spector et al., 2001; Lerner et al., 2001)(Fig. 2). The second localizer was used to identify regions of retinotopic early visual cortex (including V1, V2, V3, and V4) corresponding to the central visual field where the object images were presented. Participants viewed alternating 16 s blocks of a central disk (5 °), and an annulus (6 – 15 °). The identified region of central early visual cortex was divided into V1 (posterior) and extrastriate (anterior) retinotopic cortex (ES) using the method developed by Hinds and colleagues to identify V1 (Hinds et al., 2008)(Fig. 2).

fMRI data acquisition

Participants were scanned on the 3T General Electric scanners at the fMRI facility on the NIH campus in Bethesda. Images were acquired using an 8-channel head coil with an in-plane resolution of 2×2 mm, and 19 2 mm slices (0.2 mm inter-slice gap, repetition time [TR] = 2 s, echo time [TE] = 30 ms, matrix size = 96×96, field of view (FOV) = 192mm). Given the small voxel size, we could only acquire partial volumes of the temporal and occipital cortices. Our slices were oriented approximately parallel to the base of the temporal lobe. Thus, our data did not include the parietal or temporal lobe or regions of the auditory cortex. All functional localizer and main task runs were interleaved. During the main task, fMRI BOLD signals were measured while participants either viewed or imagined each of the 10 objects (Fig. 1).

fMRI Data Analysis

Data analysis was conducted using AFNI (http://afni.nimh.nih.gov/afni) with SUMA (AFNI surface mapper) and custom MATLAB scripts. Data preprocessing included slice-time correction, motion-correction, and smoothing (smoothing was performed only for the localizer data, not the event-related data, with Gaussian blur of 5 mm full-width half-maximum). To deconvolve the event-related responses, a standard general linear model using the AFNI software package was conducted. We examined patterns of neural activity within both early visual cortex (V1 and ES) and object-selective regions (LO and pFs). Because the results were similar between the left and right hemisphere in all 4 regions-of-interest (ROIs), we here collapsed across hemisphere. Within each of the ROIs, Multi-Voxel Pattern Analysis (MVPA) was used to assess the stimulus information available in the pattern of response.

To investigate discrimination of individual objects, we used the split-half correlation analysis method as the standard measure of information (Chan et al., 2010; Haxby et al., 2001; Kravitz et al., 2010; Reddy and Kanwisher, 2007; Williams et al., 2008). Briefly, the 6 event-related runs for each participant were divided into two halves (each containing 3 runs) in all possible 10 ways. For each of the splits, we estimated the t-value between each condition and baseline in each half of the data, and then extracted t-values from the voxels within each ROI. Before calculating the correlations, the t-values were normalized separately in each voxel for the perception and imagery conditions by subtracting the mean value across all conditions ("cocktail blank")(Haxby et al., 2001). This normalization was performed separately for perception and imagery given the large difference in the magnitude of response between these two different trial types. Specifically, the mean t-value across all perception conditions was subtracted from of the t-value of each individual perception condition, and similarly, the mean t-value across all imagery conditions was subtracted from of the individual t-value for each imagery condition.

Correlation coefficients (Pearson) were calculated by comparing the normalized t-values of an object condition with the values of every other condition.

To compare the structure of representations across ROIs, we combined data from all 6 runs. As with the split-half analysis, we subtracted the cocktail blank from all perception or imagery conditions separately. This produced similarity matrices for both perceived and imagined objects, which we then correlated within and across ROIs. Since this analysis was conducted across all 6 runs, the diagonal values of the similarity matrix were not defined and were thus excluded. For the correlation matrix between imagery and perception, we averaged across both directions in which the imagery structure of a ROI was correlated with the perception structure of the other ROI and in which the perception structure of the former ROI was correlated with the imagery structure of the latter. The resulting correlations reflect the similarity of the structure of representations for seen and imagined objects across ROIs.

We repeated our discrimination analyses for imagery and perception with multi-class classification using a linear support vector machine (SVM) approach. In these classifications, we used the LIBSVM package developed by Chang and Lin (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)(Cox and Savoy, 2003). We used a leave-one-run-out procedure with 5 runs to train the classifiers, and one run to test the classifier, iterating across all possible and training and test sets.

Multidimensional scaling (MDS) was performed based on the full matrix of imagery and perception correlations with non-metric scaling. We first calculated the mean dissimilarity matrix across participants, and then derived MDS plots. These MDS plots project the dissimilarities between conditions into physical distances in two dimensions. Thus the distance between any pair of conditions in each MDS plot represents the dissimilarity of their response patterns.

To rule out any effect of ROI size in the pattern of results we observed, we also ran all analyses with ROI size equated within each participant. Exactly the same pattern of results was observed. Similarly, to control for any non-normality in the distribution of correlation values we ran all correlation analyses using Fisher's z′ transformed values. Again, there was no effect on the pattern of results we report here.

Results

Prior to scanning, participants were trained (see Material and methods) to remember and imagine the details of 10 full color object images (Fig. 1A). These object images were chosen to be distinct both retinotopically and categorically, maximizing the potential contribution of both early visual and object-selective cortex. The training required participants to process and retain as much detail from the object images as possible, so that their later imagery of the object images would be precise.

During scanning, participants were presented with fully interleaved perception and imagery trials in an event-related design (Fig. 1B). Responses were examined in both retinotopic and high-level ROIs, localized in independent scans (see Material and methods). The foveal portion of retinotopic early visual cortex was divided into posterior (V1) and anterior (extrastriate cortex, ES) regions using a probabilistic atlas of V1 location (Hinds et al., 2008) in each participant. Object-selective cortex was divided into anterior (posterior fusiform sulcus, pFs) and posterior (lateral occipital, LO) ROIs (Fig. 2).

Response Magnitude

We first examined the average magnitude of response across all voxels and object conditions within each ROI for the perception and imagery conditions. As expected, strong responses were elicited in all ROIs during the perception conditions. During imagery, however, responses, on average, were much smaller and only significantly greater than zero in ES (Fig. 3). Further, while responses tended to be stronger in object-selective than retinotopic regions during perception, the opposite pattern was observed during imagery. A three-way ANOVA with Selectivity (retinotopic, object), Location (posterior, anterior), and Type (perception, imagery) as factors revealed a highly significant main effect of Type ($F_{1,10}$ = 146.211, p < 0.01) reflecting the stronger responses during perception. There was also a significant interaction between Type and Selectivity ($F_{1,10}$ = 6.759, p < 0.05), reflecting the different profiles of responses during imagery and perception.

Perceptual Decoding

We next used MVPA to investigate the specificity of the neural representations during perception for individual object images. Note that an absence of significant average activation across an ROI (see above) does not preclude information in the fine-grain pattern of response across the ROI (e.g. Williams et al., 2008). For each object we extracted the pattern of response in each ROI in two independent halves of the data (see Material and methods). Within- (e.g. chair-chair) and between-image (e.g. chair-clock) correlations were extracted for each of the 10 object images. For each object, between-image correlations were then averaged and subtracted from the within-image correlation to yield discrimination indices (Fig. 4A; left panel)(Haxby et al., 2001; Kravitz et al., 2010). Though discrimination indices were significantly greater than zero in all ROIs (one-sample planned t-tests, all p < 0.01), indicating significant decoding of object image identity, decoding was not equivalent across ROIs. A two-way ANOVA with Selectivity and Location as factors yielded a highly significant main effect of Selectivity ($F_{1,10}$ = 68.784, p < 0.01), and also a main effect of Location ($F_{1,10}$ = 27.127, p < 0.01). This result reflects the fact that retinotopic regions evidenced better perceptual decoding than object-selective regions, and that the posterior was better than the anterior regions for both retinotopic and object-selective cortex. Subsequent paired t-tests between ROIs revealed that discrimination in V1 was significantly greater than discrimination in LO or pFs (all p < 0.001). Further, discrimination in ES was significantly greater than discrimination in LO or pFs (all p < 0.01).

To replicate this result with another approach, we also performed multi-class classification analysis using a support vector machine (SVM) (Fig. 4B). Consistent with the discrimination indices, a series of one-sample t-tests revealed that performance was significantly greater than chance in all ROIs (Fig. 4B; left panel; all p < 0.01) and a two-way ANOVA yielded a main effect of Selectivity ($F_{1,10}$ = 16.978, p < 0.01), but no main effect of Location ($F_{1,10}$ = 0.726, p > 0.41). A series of paired t-tests revealed that performance in V1 was significantly higher than in LO or pFs (all p < 0.01), and performance in ES was also higher than in LO or pFs (all p < 0.01).

Taken together, these results show that while both early visual (V1, ES) and object-selective cortex (LO, pFs) contain information about the identity of seen objects, early visual cortex has significantly more distinct representations for each perceived object image than object-selective regions.

Imagery Decoding

We next investigated the specificity of the neural representations during imagery. Discrimination indices (Fig. 4A; right panel) were significantly positive in the ES, LO, and pFs (one-sample planned t-tests: ES, LO p < 0.05; pFs, p < 0.01) but not V1 (p = 0.11). A

two-way ANOVA with Selectivity and Location as factors revealed no significant effect (all $F_{1,10} < 1.812$, $p > 0.20$). SVM classifier performance showed a similar pattern, with weak but significantly above chance discrimination of individual objects in object-selective and ES cortex (Fig. 4B; right panel; one sample t-tests: ES, $p < 0.05$; LO, pFs, $p < 0.01$), but not in V1 ($p > 0.46$). Thus, both discrimination indices and SVM classification demonstrate that visual areas outside of V1 hold information about the identity of imagined objects.

To compare imagery and perception decoding we tested directly the differences in discrimination indices. A three-way ANOVA with Selectivity, Location, and Type as factors revealed a significant main effect of Type ($F_{1,10} = 54.113$, $p < 0.01$), reflecting the much stronger discrimination during perception than imagery. There was also a significant interaction between Type and Selectivity ($F_{1,10} = 57.727$, $p < 0.01$), reflecting the different profiles of decoding during perception and imagery: a strong posterior-anterior gradient during perception but no or even a reversed gradient during imagery. In fact, when the ability of each ROI to decode imagined objects is considered relative to the decoding for seen objects, the gradient from V1 to high-level visual cortex is even stronger (Fig. 5). Whereas discrimination during imagery is roughly 40% of discrimination during perception in object-selective regions, it is only 10% of discrimination during perception in retinotopic regions. A two-way ANOVA with Selectivity and Location as factors revealed a significant main effect of Selectivity ($F_{1,10} = 7.251$, $p < 0.05$) reflecting the greater relative imagery decoding in object-selective compared with retinotopic cortex.

Overall, these results demonstrate that the patterns of response in visual cortical areas can be used to decode the identity of both seen and imagined objects. Importantly however, the distribution of information across visual areas during imagery differs significantly from that observed during perception, suggesting that the relative roles of retinotopic and object-selective cortex during imagery and perception are reversed.

Decoding between imagery and perception

To investigate the relationship between the responses observed during imagery and perception, we first used MDS (see Material and methods) to qualitatively visualize the relationship between perception and imagery representations (Fig. 6A) before conducting more quantitative analyses of the relationship between imagery and perception (see below).

MDS takes into account the correlation between every pairing of object both within and across imagery and perception. There are two interesting patterns that emerge from this analysis. First, there is a general alignment between imagery (Fig. 6A; red backgrounds) and perception (Fig. 6A; blue backgrounds) in all visual areas, although this tends to be stronger in object-selective regions. For example, for both bag and pen, imagery and perception of each object lie close to each other in the MDS plots for all visual areas. Further, in both pFs and LO, the closest seen object to an imagined object was the matching object for four objects. Across objects, the median rank of the matching conditions between imagery and perception was smaller than chance rank in each ROI (V1: 3.5; ES: 3.0; LO: 4.5; pFs: 2.5; chance rank: 5.5).

Second, there are changes in the overall separation both within and between the perception and imagery conditions. In particular, in pFs, the seen and imagined objects are well intermixed, whereas in V1, they are largely segregated with the imagery conditions closer to the center of the space. This effect can be measured by the ratio of the average distance between the imagery and perception conditions (imagery/perception; V1: 0.56; ES: 0.61; LO: 0.78; pFs: 0.81). Thus, in V1, the imagery conditions appear closer to each other than the perception conditions, whereas in pFs, the distances within perception and imagery are

more similar. In sum, this pattern suggests that there is greater similarity between perception and imagery in the anterior object-selective regions, particularly pFs.

Next, to more quantitatively examine the correspondence between imagery and perception, we calculated discrimination indices based on the correlations between imagery and perception (e.g. the correlation between visually-presented and imagined chair minus the average correlation between visually-presented chair and all other imagined objects). These indices provide a measure of whether the representations of seen and imagined objects are similar enough to enable decoding across them. Consistent with the qualitative pattern observed in the MDS, cross-decoding was significant in all ROIs (Fig. 6B; one-sample planned t-tests: V1, ES, LO, p < 0.05; pFs, p < 0.01). SVM classifier performance also showed a similar pattern, with significantly above chance discrimination of individual objects in all ROIs (Fig. 6C; one sample t-tests: V1, LO, p < 0.05; ES, pFs, p < 0.01). Further, a two-way ANOVA on the cross-decoding discrimination indices and SVM classifier results with Selectivity and Location as factors revealed no significant effects (all $F_{1,10} < 3.126$, p > 0.10) and t-tests revealed no pairwise differences (p > 0.13), suggesting equivalent cross-decoding across all regions and ROIs. Interestingly, this result shows that it was possible to decode object image identity across perception and imagery even in the retinotopic ROIs where decoding of imagery alone did not reach significance. This result arises because of the much stronger decoding of perceived than imagined objects, making the comparison of imagery and perception more stable than the comparison of imagery with imagery.

Taken together, these results suggest that there is at least some agreement between the representations of seen and imagined objects across the ventral visual pathway. Note however, both the seen and perceived objects were preceded by an identical spoken word that might have contributed to the observed cross-decoding. However, spoken words are complex stimuli and to date there is no evidence they can be decoded from the response of primary auditory cortex, let alone early visual areas. Further, we note that if we had only presented the auditory labels during the imagery condition (as in Stokes et al, 2009, 2011), any region that showed imagery but not perceptual decoding could have been solely related to the auditory cue and not to imagery.

Comparison of the structure of representations between ROIs

We next directly compared the structure of the representations across ROIs during imagery and perception. For each ROI, there is a particular pattern of correlations between each of the 10 objects and all other objects, comprising the overall similarity matrix. This similarity matrix defines the structure of the representations of these objects for a particular combination of ROI and Type. These matrices can be directly compared. For example, in pFs and LO, the structure of representations was very similar during perception (Fig. 7A). In particular, in both ROIs, 'clock' showed a high correlation with 'violin' but a much lower correlation with 'chair'. Similarly, in both ROIs, 'lamp' showed a much higher correlation with 'umbrella' than with 'clock'.

It is important to note that the discrimination analysis we conducted earlier relies on measuring how replicable the pattern of response within a region is across independent presentations of seen/imagined objects relative to the similarity of the response pattern to different seen/imagined objects. However, the structure analysis is taken across the entire data set and evaluates how the structure of the representations correlates across independent regions during, essentially, a single presentation of each seen/imagined object.

To systematically compare the structure of representations for imagery and perception across ROIs, we cross-correlated the similarity matrices (Fig. 7B). This revealed significant

correlations between all ROIs during both perception and imagery (all p < 0.01) with stronger correlations during imagery than during perception (see below). Comparing between imagery and perception (Fig. 7B, right panel) revealed significant correlations within ROI for LO and pFs (all p < 0.05), but not V1 and ES (p > 0.15). Thus, consistent with our earlier analyses, the representations during perception and imagery tend to be more similar in higher than lower visual areas. The structure of representations between imagery and perception was also correlated across ROIs, with significant correlations for all comparisons excluding V1 (all p < 0.05).

MDS performed on the correlations between ROIs (Fig. 7C) revealed a strong separation between perception (blue labels) and imagery (red labels) with greater separation between the ROIs under perception than under imagery. To investigate the separation within perception or imagery more closely, we first focused on the correlation values between V1 and the other ROIs (ES, LO, or pFs) (Fig. 7D, left panel). In both perception and imagery, there was a posterior to anterior gradient with ES showing the highest correlation and pFs showing the lowest correlation with V1. Moreover, the correlation between the ROIs during visual imagery was significantly higher than the correlation during perception. A two-way ANOVA with ROI (ES, LO, pFs) and Type (Imagery, Perception) as factors revealed significant main effect of Type ($F_{1,10} = 7.909$, p < 0.05), arising from the higher correlations during imagery, and also of ROI ($F_{2,20} = 37.477$, p < 0.01), reflecting the weaker correlations with increasing separation of the ROIs.

Next, we examined the correlation between pFs and the other ROIs (V1, ES, or LO) (Fig. 7D, right panel). We observed the opposite gradient to that observed with V1 above. The highest correlation was with LO and lowest with V1. As above, the correlation between the ROIs during imagery was significantly higher than the correlation during perception. These two effects led to significant main effects of Type ($F_{1,10} = 11.299$, p < 0.01) and ROI ($F_{2,20} = 25.607$, p < 0.01) again reflecting the stronger correlations during imagery than perception and the weaker correlations with increasing separation of the ROIs.

These results demonstrate a striking difference between imagery and perception: the overall structure of representations between ROIs is much more similar during imagery than during perception. This difference likely reflects differences in the neural dynamics operating during perception and imagery (see Discussion).

Correlation with Behavior

As part of the pre-scan imagery training, we asked participants to complete a standardized battery of visualization questions (VVIQ)(Cui et al., 2007; Marks, 1973), and previous research has reported a correlation between this VVIQ measure and measures of brain activity during imagery (Amedi et al., 2005; Cui et al., 2007). VVIQ is a measure of the subjective vividness of mental images, which is akin to gauging the similarity between mental and actual image, making cross-decoding between imagery and perception the ideal imaging measure for comparison. We investigated whether the correspondence between imagery and perception was correlated with VVIQ. To increase power we combined the two early visual and object-selective ROIs into two larger ROIs (retinotopic and object-selective). We extracted the average discrimination index across imagery and perception in early visual and object-selective cortex for each participant and correlated these with VVIQ scores (Fig. 8). In both retinotopic and object-selective regions, there was a positive correlation between discrimination indices and VVIQ, although this correlation was only significant in early visual areas (r = 0.691, p < 0.01) and not in object-selective cortex (r = 0.378, p = 0.126). VVIQ was not significantly correlated with pure imagery decoding in either EVC (r = 0.007, p = 0.49) or object-selective cortex (r = 0.329, p = 0.162). Thus, only

our measure of the similarity between perception and imagery reflects subjective measures of the strength of imagery, especially in early visual cortex.

Discussion

In the present study, we used fMRI to conduct a detailed comparison of visual representations during imagery and perception. First, we examined the *specificity* of representations, and found that during both perception and imagery, the patterns of response in visual areas could be used to decode an individual seen or imagined object (out of 10 possible object images), though decoding was much weaker during imagery than perception. Further, there was enough correspondence between imagery and perception to allow cross-decoding of objects throughout the ventral visual pathway. Second, we investigated the *distribution* of perception and imagery information. While there was a strong posterior-anterior gradient across ROIs in the strength of seen object decoding, this gradient was absent or even reversed for imagined objects. Third, we investigated the *differences* between the perceptual and imagery representations of individual objects. Direct comparison of the structure of representations between ROIs during perception and imagery found greater similarity across ROIs during imagery than perception. Finally, we found that cross-decoding between perception and imagery correlated with a subjective measure of the vividness of visual imagery, particularly in retinotopic cortex. These results highlight that while imagery and perception may involve the same areas of visual cortex, even at a fine grain of analysis, there are still striking differences in the response of those areas in imagery and perception.

Distinct coding of individual objects during visual imagery

Our results clearly demonstrate that there is specific pattern of neural activation throughout visual areas for single objects during both visual imagery and perception, and that these patterns are related (Fig. 4 and Fig. 6). These findings significantly extend prior work showing category-specific activation in high-level visual areas during imagery (Ishai et al., 2000; O'Craven and Kanwisher, 2000; Reddy et al., 2010) or decoding of two simple imagined stimuli (the letters 'X' and 'O')(Stokes et al., 2009). First, we conducted a rigorous test of the specificity of visual imagery representations, by using a set of 10 possible object images, none corresponding to categories (e.g. faces, houses) for which category-selective regions have been identified in cortex (cf. Reddy et al., 2010). Second, we systematically compared the distribution of visual imagery representations across different regions of the ventral visual pathway, including V1 and extrastriate retinotopic cortex. Such a detailed comparison across the ventral visual pathway has not been provided by previous studies, which often focused on specific visual areas using stimuli or comparisons tailored for that area (e.g. Gabor filters in V1, Thirion et al. 2006; categories in object-selective cortex, Reddy et al., 2010). In many cases, perceptual decoding was not present in all visual areas making it impossible to compare the relation between perception and imagery across areas (e.g. Stokes et al. 2009).

While our findings are consistent with those showing some decoding of an 'X' and an 'O' during imagery, as well as some evidence for cross-decoding between imagery and perception (Stokes et al., 2009), there are a number of discrepancies between the findings of the two studies. In particular, while Stokes and colleagues found evidence for some cross-decoding in left pFs only, we find significant decoding between imagery and perception in all visual areas (V1, ES, LO, pFs). Further, Stokes and colleagues found little difference between imagery and perception in visual areas either in terms of decoding or activation. In fact they found that decoding was more reliable during imagery than during perception especially in pFs. In contrast, we found better discrimination during perception than imagery and much stronger activation for perception than imagery throughout visual areas. These

differences between our findings and those of Stokes and colleagues may reflect the fact that in their study i) auditory cues were only provided during imagery and not perception trials, ii) perceptual stimuli were only presented for 250 ms whereas imagery is likely to have last much longer, iii) imagery and perception trials were blocked rather than fully interleaved as in our design, or iv) simple letter stimuli may not be ideally suited for probing representations in object-selective cortex (Joseph et al., 2003; Sergent et al., 1992).

Relationship between imagery and perception

While we found similarities in the representation of individual objects during imagery and perception, the two processes were also distinct, suggesting that imagery is not just a weak form of perception. This dissociation between imagery and perception was revealed from the detailed comparison of results across multiple visual areas (Fig. 4 and Fig. 7). In particular, we found that discrimination during imagery and between perception and imagery was similar for all visual areas, whereas there was a strong gradient of decoding during perception from V1 to pFs. This difference in gradients is similar to that previously reported for another top-down process, "refreshing" (Johnson et al., 2007). Further, the structure of representations across ROIs was more similar during imagery than perception (Fig. 7). These findings significantly extend a prior report suggesting that visual perception and imagery are distinguishable by deactivation of auditory cortex during imagery but not perception (Amedi et al., 2005). Here we show that imagery and perception are distinct even within visual modality processing.

The differences we find between imagery and perception likely reflect differences in network dynamics during imagery and perception. During visual perception, all visual information about perceived stimuli is available at the level of the retina and each visual region transforms this representation to highlight particular aspects of that information (DiCarlo and Cox, 2007; Kravitz et al., in press). This unique representation is produced in each region based on three basic dynamics: 1) bottom-up input carrying perceptual information, 2) top-down modulatory feedback, and 3) internal processing that integrates the bottom-up and top-down signals to produce a unique representation. During imagery, bottom-up input is absent, leaving each region with only top-down feedback as input. This reduction in information likely has two effects. First, it reduces the complexity of the internal processing within any given region, and, second, because of this reduced processing there is likely less transformation of the signal between regions. These effects will lead individual regions to have less unique representations. Further, the top-down signal is undoubtedly impoverished in its information relative to the full detail available in the actual images of the objects. Therefore, the information critical for generating a unique representation in any particular region during perception is likely reduced or even unavailable in the top-down signal. In general, the removal of bottom-up input will alter the dynamics of the internal processing within each region and the communication among regions, leading to increased correlation in the representations between regions.

This model suggests that while imagery and perception depend on similar neural substrates (Fig. 6), the contributions of the substrates to each process differ (Figs. 4 and 7). This provides a resolution to the long-standing conflict between findings suggesting the same regions are activated during imagery and perception and findings of a double dissociation between these processes. Double dissociations are generally taken as evidence of non-overlapping neural substrates, but damage to the same general area can differentially impair processes that use the same substrate depending on how much it interrupts the processing critical to either (Plaut, 1995) (Plaut and Shallice, 1993). Our finding of increased correlation between regions during imagery provides evidence of differential processing in imagery and perception within the same neural substrate, suggesting that lesions of the same area could selectively impair either.

Visual imagery representations in early visual cortex

It has been a long-standing debate whether early visual cortex, a low level visual structure, is activated during visual imagery (Kosslyn et al., 2001). This debate is also intertwined with the nature of visual imagery. One class of imagery theories argues that visual imagery relies on depictive (picture-like) representations (Kosslyn and Thompson, 2003; Slotnick et al., 2005) whereas the other class of theories posits that visual imagery relies entirely on symbolic (language-like) representations (Pylyshyn, 2002). Thus, only the depictive view predicts a depictive role for early visual cortex in visual imagery. We found significant decoding in ES for imagery (Fig. 4) and cross-decoding between imagery and perception in both V1 and ES (Fig. 6B). These decoding results support the depictive view of imagery (Kosslyn and Thompson, 2003) by showing decoding of imagined objects in early visual cortex as well as reduced or absent cross-decoding with less vivid imagery. Our results are consistent with prior reports of decoding in early visual cortex during imagery (Thirion et al., 2006) and in the context of higher-level processing such as semantic knowledge or working memory (Harrison and Tong, 2009; Hsieh et al., 2010; Kosslyn et al., 1995; Serences et al., 2009). However, V1 showed only weak, non-significant decoding during visual imagery (Fig. 4), and when the ability of early visual cortex to decode seen objects is considered, the weakness of imagery decoding is even more striking (Fig. 5). Overall, these results indicate that while early visual cortex does contain signals during visual imagery, its relative contribution during imagery of complex objects compared with other visual areas is much weaker than during perception of those same objects. This is not to suggest that early visual cortex does not play a critical and perhaps even central role in the imagery of other simpler visual stimuli (e.g. oriented lines or highly distinct retinotopic patterns)(Thirion et al., 2006). In fact, our behavioral result (Fig. 8) suggests that early visual cortex has a pivotal role in generating visual mental images of even complex visual stimuli.

Conclusions

In summary, our findings show that the relationship between imagery and perception is complex, with overlap in the neural substrates involved, but differences in how that tissue is utilized. We suggest that these differences directly result from the removal of bottom-up input during imagery and changes in the neural dynamics across regions.

Acknowledgments

References

Amedi A, Malach R, Pascual-Leone A. Negative BOLD differentiates visual imagery and perception. Neuron. 2005; 48:859–872. [PubMed: 16337922]

Bartolomeo P. The relationship between visual perception and visual mental imagery: a reappraisal of the neuropsychological evidence. Cortex. 2002; 38:357–378. [PubMed: 12146661]

Bartolomeo P. The neural correlates of visual mental imagery: an ongoing debate. Cortex. 2008; 44:107–108. [PubMed: 18387539]

Behrmann M. The mind's eye mapped onto the brain's matter. Curr Psychol Science. 2000; 9:50–54.

Behrmann M, Moscovitch M, Winocur G. Intact visual imagery and impaired visual perception in a patient with visual agnosia. J Exp Psychol Hum Percept Perform. 1994; 20:1068–1087. [PubMed: 7964528]

Behrmann M, Winocur G, Moscovitch M. Dissociation between mental imagery and object recognition in a brain-damaged patient. Nature. 1992; 359:636–637. [PubMed: 1406994]

Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI. Cortical representations of bodies and faces are strongest in commonly experienced configurations. Nat Neurosci. 2010; 13:417–418. [PubMed: 20208528]

Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage. 2003; 19:261–270. [PubMed: 12814577]

Cui X, Jeter CB, Yang D, Montague PR, Eagleman DM. Vividness of mental imagery: individual variability can be measured objectively. Vision Res. 2007; 47:474–478. [PubMed: 17239915]

DiCarlo JJ, Cox DD. Untangling invariant object recognition. Trends Cogn Sci. 2007; 11:333–341. [PubMed: 17631409]

Farah, MJ. Mental imagery. In: Gazzaniga, M., editor. The cognitive neurosciences. Vol. 2. Cambridge, MA: MIT Press; 1999. p. 965-974.

Farah MJ, Levine DN, Calvanio R. A case study of a mental imagery deficit. Brain and Cognition. 1988; 8:147–164. [PubMed: 3196480]

Ganis G, Thompson WL, Kosslyn SM. Brain areas underlying visual mental imagery and visual perception: an fMRI study. Brain Res Cogn Brain Res. 2004; 20:226–241. [PubMed: 15183394]

Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. Vision Res. 2001; 41:1409–1422. [PubMed: 11322983]

Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009; 458:632–635. [PubMed: 19225460]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001; 293:2425–2430. [PubMed: 11577229]

Hinds OP, Rajendran N, Polimeni JR, Augustinack JC, Wiggins G, Wald LL, Diana Rosas H, Potthast A, Schwartz EL, Fischl B. Accurate prediction of V1 location from cortical folds in a surface coordinate system. Neuroimage. 2008; 39:1585–1599. [PubMed: 18055222]

Hsieh PJ, Vul E, Kanwisher N. Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. J Neurophysiol. 2010; 103:1501–1507. [PubMed: 20071627]

Ishai A, Sagi D. Common mechanisms of visual imagery and perception. Science. 1995; 268:1772–1774. [PubMed: 7792605]

Ishai A, Ungerleider LG, Haxby JV. Distributed neural systems for the generation of visual images. Neuron. 2000; 28:979–990. [PubMed: 11163281]

Johnson MR, Mitchell KJ, Raye CL, D'Esposito M, Johnson MK. A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. Neuroimage. 2007; 37:290–299. [PubMed: 17574442]

Joseph JE, Gathers AD, Piper GA. Shared and dissociated cortical regions for object and letter processing. Brain Res Cogn Brain Res. 2003; 17:56–67. [PubMed: 12763192]

Kastner S, De Weerd P, Desimone R, Ungerleider LG. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science. 1998; 282:108–111. [PubMed: 9756472]

Klein I, Dubois J, Mangin JF, Kherif F, Flandin G, Poline JB, Denis M, Kosslyn SM, Le Bihan D. Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. Brain Res Cogn Brain Res. 2004; 22:26–31. [PubMed: 15561497]

Kosslyn SM, Ganis G, Thompson WL. Neural foundations of imagery. Nat Rev Neurosci. 2001; 2:635–642. [PubMed: 11533731]

Kosslyn SM, Thompson WL. When is early visual cortex activated during visual mental imagery? Psychol Bull. 2003; 129:723–746. [PubMed: 12956541]

Kosslyn SM, Thompson WL, Alpert NM. Neural systems shared by visual imagery and visual perception: a positron emission tomography study. Neuroimage. 1997; 6:320–334. [PubMed: 9417974]

Kosslyn SM, Thompson WL, Kim IJ, Alpert NM. Topographical representations of mental images in primary visual cortex. Nature. 1995; 378:496–498. [PubMed: 7477406]

Kourtzi Z, Kanwisher N. Cortical regions involved in perceiving object shape. J Neurosci. 2000; 20:3310–3318. [PubMed: 10777794]

Kravitz, DJ.; Chan, AW-Y.; Baker, CI. Investigating high-level visual representations: objects, bodies, and scenes. In: Kriegeskorte, N.; Kreiman, G., editors. Visual Population Codes. 2011.

Kravitz DJ, Kriegeskorte N, Baker CI. High-Level Visual Object Representations Are Constrained by Position. Cereb Cortex. 2010; 12:2916–2925. [PubMed: 20351021]

Kreiman G, Koch C, Fried I. Imagery neurons in the human brain. Nature. 2000; 408:357–361. [PubMed: 11099042]

Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R. A hierarchical axis of object processing stages in the human visual cortex. Cereb Cortex. 2001; 11:287–297. [PubMed: 11278192]

Marks DF. Visual imagery differences in the recall of pictures. Br J Psychol. 1973; 64:17–24. [PubMed: 4742442]

Moro V, Berlucchi G, Lerch J, Tomaiuolo F, Aglioti SM. Selective deficit of mental visual imagery with intact primary visual cortex and visual perception. Cortex. 2008; 44:109–118. [PubMed: 18387540]

O'Craven KM, Kanwisher N. Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. J Cogn Neurosci. 2000; 12:1013–1023. [PubMed: 11177421]

Pearson J, Clifford CW, Tong F. The functional impact of mental imagery on conscious perception. Curr Biol. 2008; 18:982–986. [PubMed: 18583132]

Plaut DC. Double dissociation without modularity: evidence from connectionist neuropsychology. J Clin Exp Neuropsychol. 1995; 17:291–321. [PubMed: 7629273]

Plaut DC, Shallice T. Deep dyslexia: a case study of connectionist neuropsychology. Cognitive Neuropsychol. 1993; 10:377–500.

Pylyshyn ZW. Mental imagery: in search of a theory. Behav Brain Sci. 2002; 25:157–182. discussion 182–237. [PubMed: 12744144]

Reddy L, Kanwisher N. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol. 2007; 17:2067–2072. [PubMed: 17997310]

Reddy L, Tsuchiya N, Serre T. Reading the mind's eye: decoding category information during mental imagery. Neuroimage. 2010; 50:818–825. [PubMed: 20004247]

Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-specific delay activity in human primary visual cortex. Psychol Sci. 2009; 20:207–214. [PubMed: 19170936]

Sergent J, Zuck E, Levesque M, MacDonald B. Positron emission tomography study of letter and object processing: empirical findings and methodological considerations. Cereb Cortex. 1992; 2:68–80. [PubMed: 1633409]

Slotnick SD, Thompson WL, Kosslyn SM. Visual mental imagery induces retinotopically organized activation of early visual areas. Cereb Cortex. 2005; 15:1570–1583. [PubMed: 15689519]

Stokes M, Saraiva A, Rohenkohl G, Nobre AC. Imagery for shapes activates position-invariant representations in human visual cortex. Neuroimage. 2011; 56:1540–1545. [PubMed: 21376815]

Stokes M, Thompson R, Cusack R, Duncan J. Top-down activation of shape-specific population codes in visual cortex during mental imagery. J Neurosci. 2009; 29:1565–1567. [PubMed: 19193903]

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S. Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage. 2006; 33:1104–1116. [PubMed: 17029988]

Williams MA, Baker CI, Op de Beeck HP, Shim WM, Dang S, Triantafyllou C, Kanwisher N. Feedback of visual object information to foveal retinotopic cortex. Nat Neurosci. 2008; 11:1439–1445. [PubMed: 18978780]

Winawer J, Huk AC, Boroditsky L. A motion aftereffect from visual imagery of motion. Cognition. 2010; 114:276–284. [PubMed: 19853246]

Highlights

- Both scene and imagined object identity can be decoded.

- Information was differentially distributed for imagined and seen objects.

- The structure of representations was more similar during imagery than perception

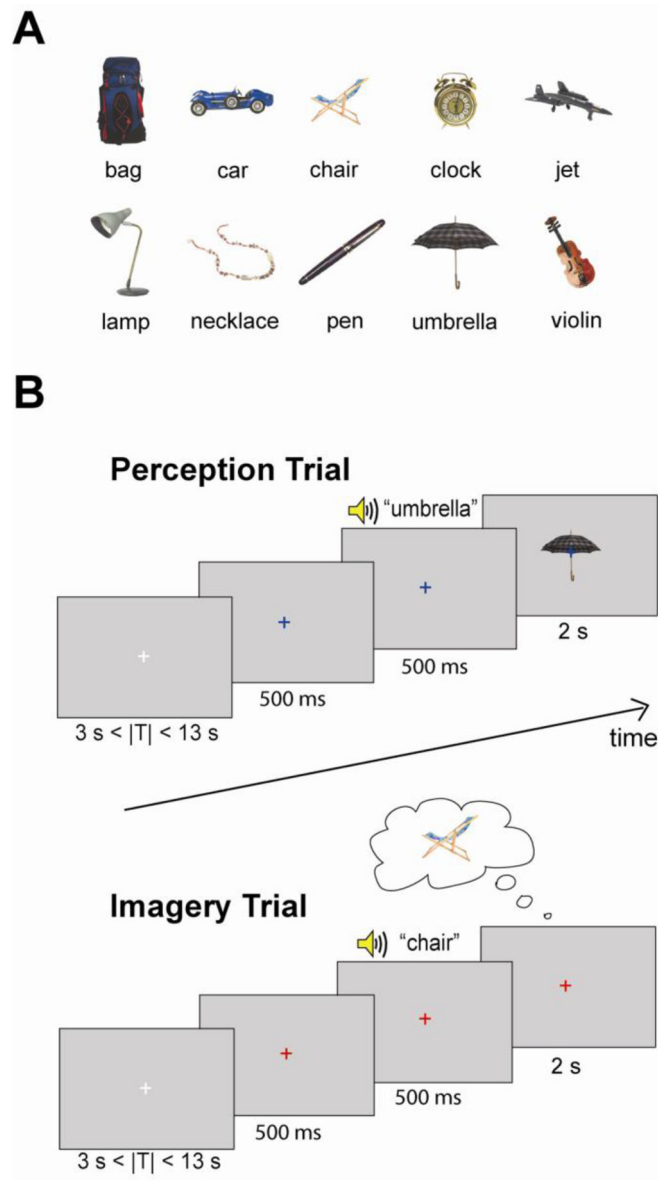- Imagery and perception involve different dynamics across the ventral visual pathway.

Figure 1.
A, 10 object images used in the main task. B, Experimental design. The main task consisted of two interleaved trial types: imagery and perception. On each trial, the white fixation cross first changed color indicating either a perception trial (blue) or an imagery trial (red) and was followed by the auditory presentation of the names of one of the objects. In perception trials, participants were instructed to view the visually-presented object. In imagery trials, participants were asked to imagine the object image given corresponding to the auditory signal. The inter-trial interval (ITI) was randomized between 3 and 13 s.
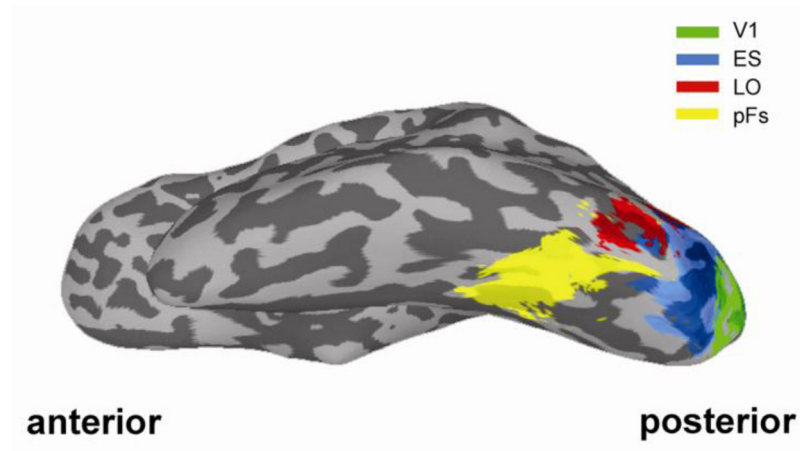
Figure 2.
Regions-of-interest (ROIs). For illustrative purposes, ROIs derived in each participant were transformed to a standard space. Green (V1), blue (ES), red (LO), or yellow (pFs) areas indicate voxels which were with the ROI in at least 3/11 subjects (27 %).
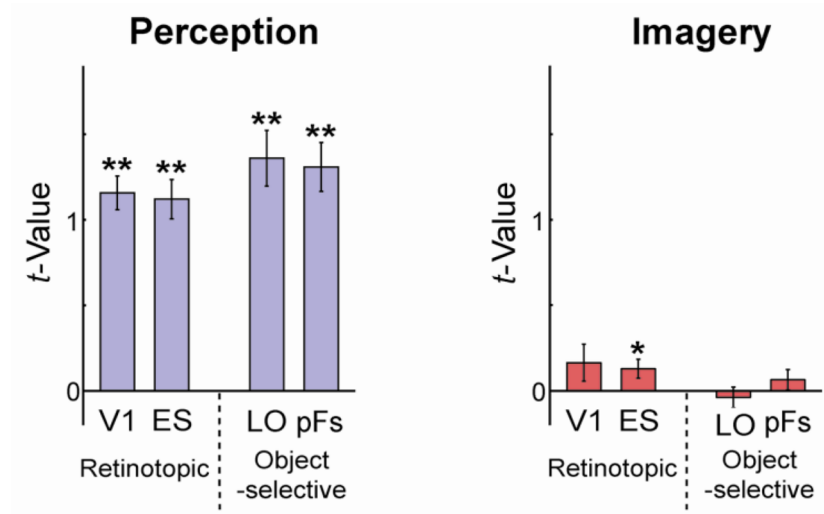
Figure 3.
The average magnitude of response in retinotopic (V1, ES) and object-selective (LO and pFs) regions during perception and imagery. Response was roughly 10 times larger during perception than imagery. Error bars represent ± 1 SEM, calculated across participants (for one-sample planned t-tests, *: $p < 0.05$; **: $p < 0.01$).
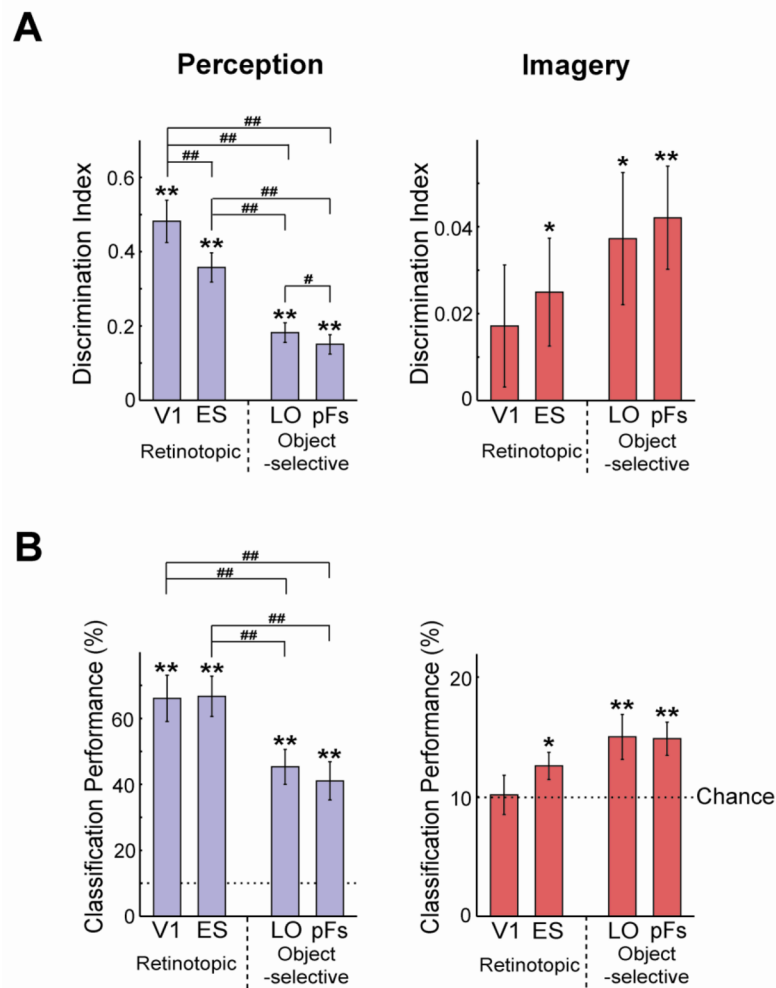
Figure 4.
The specificity of object information during imagery and perception in retinotopic and object-selective regions. A, Average discrimination indices across 10 objects during perception (left panel; blue bars) and imagery (right panel; red bars). Discrimination was strong during perception with a prominent posterior to anterior gradient. During imagery, discrimination was significant in ES, LO, or pFs. Notably, however, there was either no or a reverse gradient compared with perception. B, Classification performance with SVMs showed a similar pattern to the discrimination indices. The chance level was 10%. All error bars represent ± 1 SEM, calculated across participants (for one-sample planned t-tests, *: $p <$ 0.05; **: $p < 0.01$) (for paired t-tests, #: $p < 0.05$; ##: $p < 0.01$).
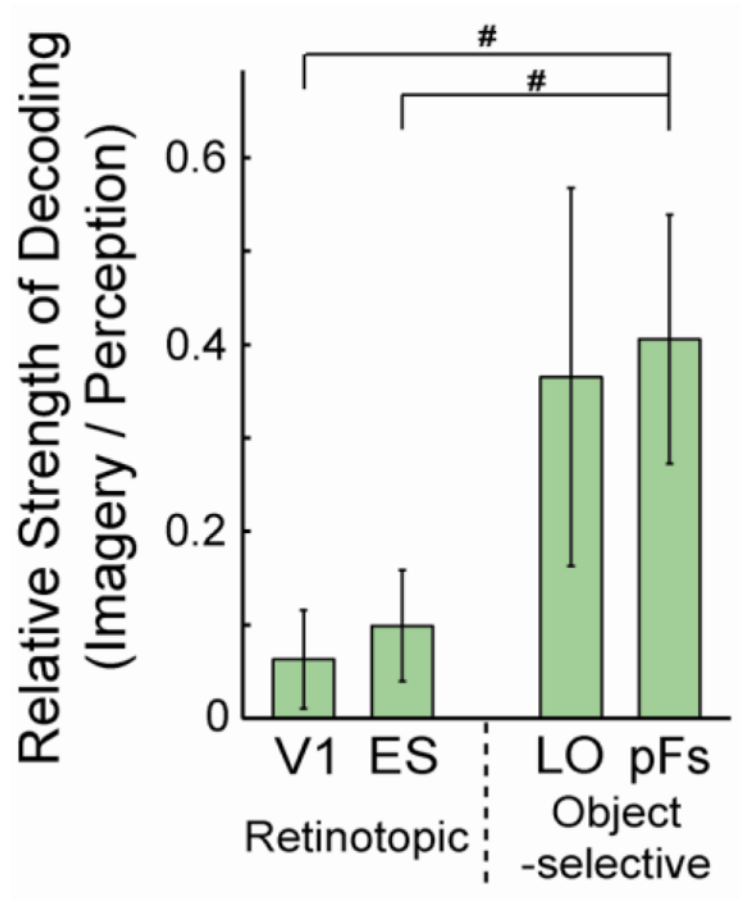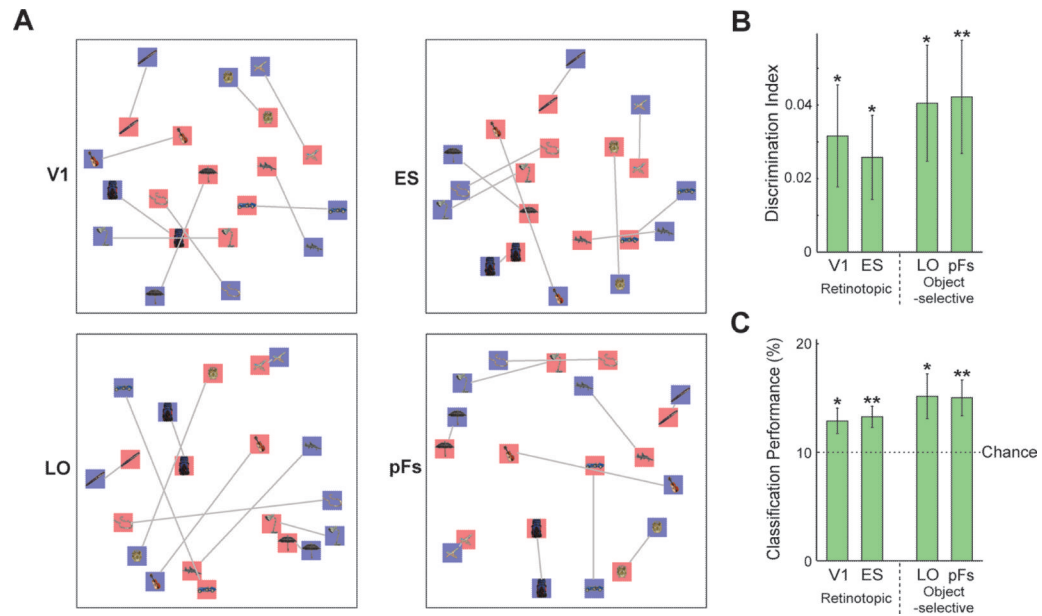
Figure 5.
Relative strength of decoding: imagery to perception. The ratio of the discrimination indices during imagery to the indices during perception was significantly greater in object-selective regions than retinotopic regions. All error bars represent ± 1 SEM, calculated across participants (for paired t-tests, #: p < 0.05).

Figure 6.
Similarity between visual imagery and perception. A, Relations between object conditions based on Multi Dimensional Scaling (MDS). Blue and red backgrounds indicate perception and imagery conditions, respectively, and gray lines indicate the matching object conditions. Imagery and perception conditions aligned, particularly in pFs. The MDS also highlights the greater similarity between imagery conditions compared with perception conditions in V1, much less than in pFs. B, Discrimination indices between seen and the corresponding imagined objects in V1, ES, LO and pFs. Discrimination of objects between imagery and perception was significant in all ROIs.. C, Classification performance with SVMs showed a similar pattern to the discrimination indices between imagery and perception conditions. The chance level was 10 %. All error bars represent ± 1 SEM, calculated across participants (for one-sample planned t-tests, *: p < 0.05; **: p < 0.01).
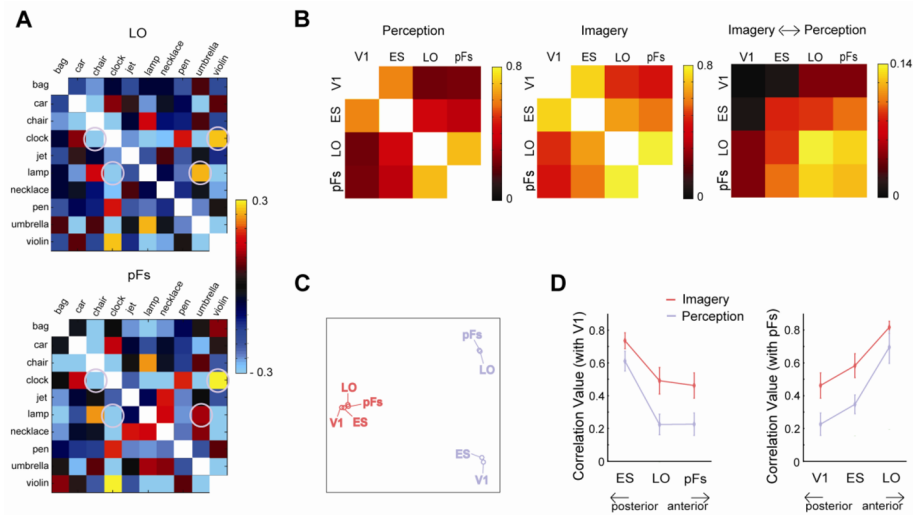
Figure 7.
Comparison of the structure of representations between ROIs across imagery and perception. A, Similarity matrices of LO and pFs during perception in a participant. Note the similar patterns of correlations between the two ROIs. B, Correlations between similarity matrices of ROIs. Correlation between ROIs during either imagery or perception tended to be high, but correlations between imagery and perception were weak. Note the different scale of correlations between imagery and perception from the cases of imagery or perception. Each off-diagonal element in the far right panel of B is the average of the correlation between the imagery matrices in one ROI (R1) and the perception matrices in the other ROI (R2): (Correlation(R1:Imagery,R2:Perception) + Correlation(R1:Perception,R2:Imagery))/2. C, The relationships between ROIs for both perception (blue labels) and imagery (red labels) based on MDS. Note the large separation between perception and imagery, and the larger separation within perception than within imagery. D, Correlations between V1 and the other ROIs (ES, LO, or pFs), and between pFs and the other ROIs (V1, ES, or LO), highlighting the stronger correlations during imagery than perception. All error bars represent ± 1 SEM, calculated across participants.
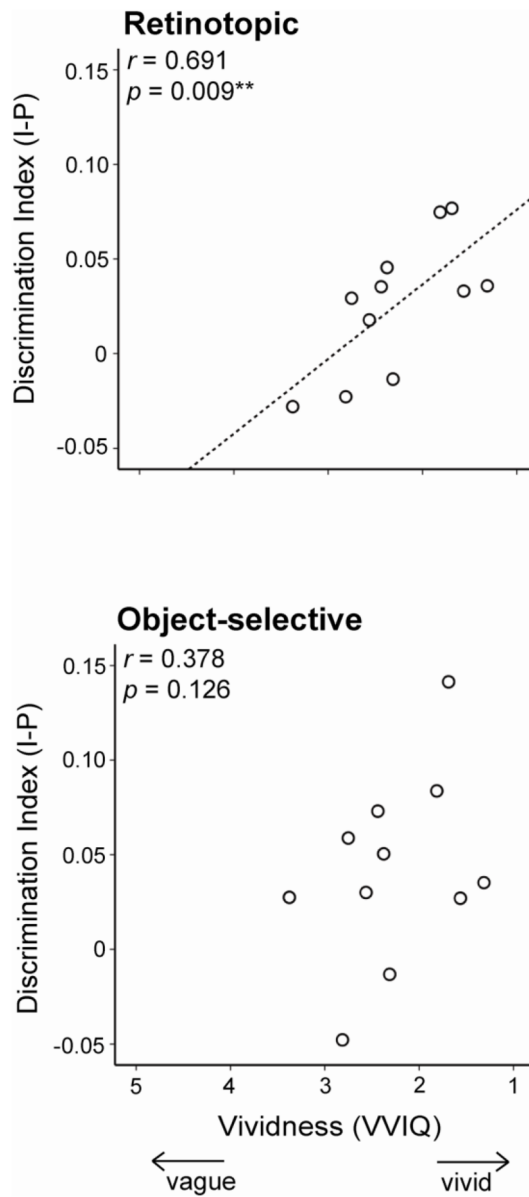
Figure 8.
Correlation between fMRI and a standardized questionnaire assessing the vividness of imagery (VVIQ). In both object-selective (lower panel) and early visual (upper panel) cortex our measure of discrimination across imagery and perception correlated with the behavioral measure of the vividness of imagery. However, this correlation was only significant in early visual cortex.