Large-scale dissociations between views of objects, scenes, and reachable-scale environments in visual cortex

Emilie L. Josephs [1] **and Talia Konkle**[1]

[a]Department of Psychology, Harvard University, 33 Kirkland St. Cambridge, MA, 02138

Space-related processing recruits a network of brain regions separate from those recruited in object processing. This dissociation has largely been explored by contrasting views of navigable-scale spaces to views of close-up, isolated objects. However, in naturalistic visual experience, we encounter spaces intermediate to these extremes, like the tops of desks and kitchen counters, which are not navigable but typically contain multiple objects. How are such reachable-scale views represented in the brain? In two human functional neuroimaging experiments, we find evidence for a large-scale dissociation of reachable-scale views from both navigable scene views and close-up object views. Three brain regions were identified which showed a systematic response preference to reachable views, located in the posterior collateral sulcus, the inferior parietal sulcus, and superior parietal lobule. Subsequent analyses suggest that these three regions may be especially sensitive to the presence of multiple objects. Further, in all classic scene and object regions, reachable-scale views dissociated from both objects and scenes with an intermediate response magnitude. Taken together, these results establish that reachable-scale environments have a distinct representational signature from both scene and object views in visual cortex.

objects | scenes | reachspaces | fMRI | visual cortex

Fig. 1. Experiment 1 stimuli and results. A) Examples of object, reachspace and scene views. B) Preference mapping results. Colored regions have preference for either objects (yellow), reachspaces (blue) and scenes (green). Color saturation indicates the magnitude of the preference relative to the next most preferred category.

Scene-based and object-based representations form a major joint in the organization of the visual system. Scene-selective brain regions are broadly concerned with performing global perceptual analysis of a space (1) (2) (3) (4), computing its navigational affordances (5) (6), and linking the present view to stored memory about the overall location (7) (8). In contrast, object-selective regions represent bounded entities, robust to confounding low-level contours and minor changes in size or position (9) (10). Are these two systems, one for processing spatial-layout and another for bounded-objects, together sufficient to represent any view of the physical environment?

Consider views of reachable-scale environments—the countertops where we combine ingredients for a cake, or the worktables where we assemble the components of a circuit board. These views are intermediate in scale to scenes and objects, and are the locus of many everyday actions (Figure 1a). How are they represented in the visual system?

One possibility is that reachable-scale environments are represented similarly to navigable-scale scenes, driving similar activations across the ventral and dorsal streams. Views of reachable environments are spatially extended, have 3D layout, and need to be situated within larger environments, all of which are hypothesized functions of scene-selective regions. However, everyday views of reachable-scale environments also prominently feature colle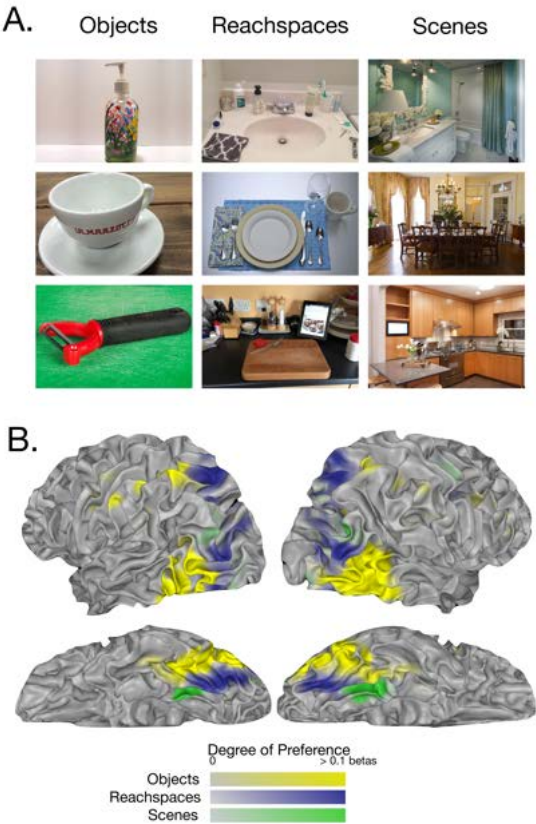ctions of multiple objects, and differ meaningfully from scenes by affording object-centered actions rather than navigation. Thus, a second possibility is that reachable-scale views will strongly drive object-preferring cortex.

A final, and not mutually-exclusive possibility, is that visual responses to reachable-scale environments might recruit distinct brain regions, separate from object- and scene- preferring cortex. There are both action-related and perception-related arguments for this hypothesis. First, it is clear that near-scale
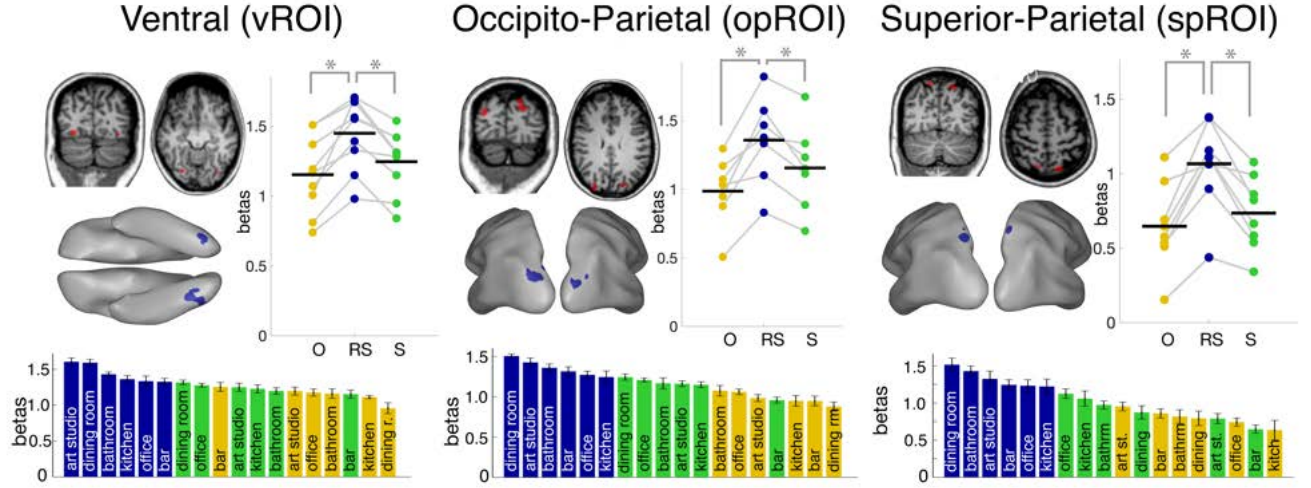
Fig. 2. Locations and activations of reachspace-preferring ROIs. ROI locations are shown in the volume and on the inflated surface of an example subject. Bar plots show beta activations for objects, reachspaces, and scenes, averaged over semantic category (3-bar plot), or with semantic category displayed separately (18-bar plot). Error bars represent the within-subject standard error of the mean, and stars indicate statistical significance.

spaces have different behavioral demands than far-scale spaces (11) (12) (13). Indeed, there are well-known motor dissociations between reach-related fronto-parietal circuits versus navigation-related medial networks (14) (15) (16). Second, statistics of visual images differ as a function of environment scale (17). We recently showed that the human perceptual system is sensitive to these differences: observers performing a visual search task were faster to find an image of a reachable environment among distractor scenes or objects than among reachspaces, and vice versa (18). These results show that the scale of the depicted environment is a major factor in perceptual similarity computations.

These prior studies suggest that reachable-scale views dissociate from singleton objects views and navigable-scale scene views in both their input-related image statistics and output-related action requirements. Such input and output pressures have been proposed to be jointly essential for the large-scale functional clustering observed in visual cortex for different kinds of visual domains (e.g. faces, scenes (19) (20) (21)(22) (23). Thus, it is possible that views of reachable environments are distinct enough in form and purpose to require distinct visual processing regions.

In the present work, we examined how views of reachable-scale environments are represented in the human brain, using functional magnetic resonance imaging. We find clear evidence that reachspace representations dissociate from those of scenes and objects. Specifically, views of reachable environments elicited greater activity than both scenes and objects in regions of ventral and dorsal occipitoparietal cortex, across variations in luminance and global spatial frequency, and variations in the semantic category depicted (e.g. kitchen vs office reachspaces). Reachable-scale environments also elicited differential responses in classic object- and scene-preferring regions, generally leading to intermediate levels of activation between scene and object views. Regions preferring reachable-scale environments showed a peripheral eccentricity bias but also responded particularly strongly to images of multiple objects, a functional signature that is distinct from both scene and object regions. Taken together, these results suggest that the

visual processing of near-scale environments is functionally dissociable from that of objects and scenes.

Results

Preferential responses to reachable-scale spaces in visual cortex. To examine the neural representation of reachable-scale environments compared to navigable-scale scenes and singleton objects, we created a stimulus set with images from each of the three environment types (see Figure 1a; Supplementary Figure 1). Object images depicted close-scale views of single objects (within 8-12 inches) on their natural background. Reachable-scale images, which we will refer to as "reachspaces", depicted near-scale environments that were approximately as deep as arm's reach (3-4ft), and consisted of multiple small objects arrayed on a horizontal surface (18). Scene images depicted views of the interior of rooms. Images were drawn from 6 semantic categories (bar, bathroom, dining room, kitchen, office, art studio). Note that we use the term "environment scale" to refer to the distinction between conditions, but caution the reader against interpreting our results in terms of subjective distance only. Rather, differences observed here likely reflect differences across a constellation of dimensions that co-occur with scale (e.g. number of objects, number of surfaces, action affordances, perceived reachability). Two stimulus sets were collected, with 90 images each (ImageSetA, ImageSetB; 30 images per environmental scale per set; see In-text Methods).

In Experiment 1, twelve participants viewed images of objects, reachspaces, and scenes, in a standard blocked fMRI design. All three stimulus conditions drove strong activations throughout visually-responsive cortex, particularly in early visual and posterior inferotemporal regions, with progressively weaker responses anteriorly through the ventral and dorsal stream (see Supplementary Figure 2). To help visualize the differences between these response topographies, voxels were colored according to the condition that most strongly activated them, with the saturation of the color reflecting the strength of the response preference (early visual regions excluded; see Supplementary Methods). This analysis revealed that different
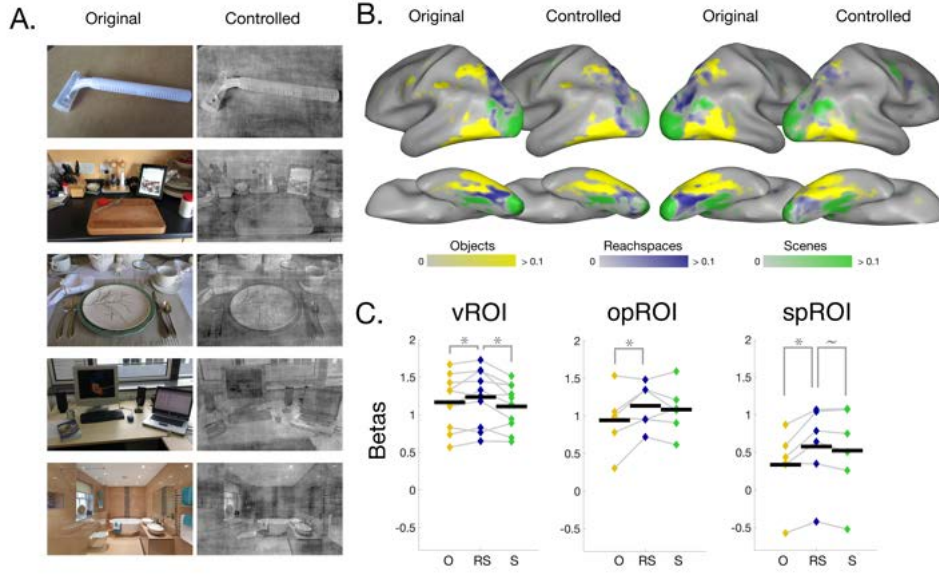
Fig. 3. Stimuli and results for Experiment 2. A) Illustration of matching luminance, contrast, and global spatial frequency. B) A comparison of the group-average preference maps obtained for the original and controlled images, plotted on the same scale and projected onto an inflated brain. Color saturation indicates the magnitude of the preference relative to the next most preferred category. C) Activations in reachspace ROIs (defined in original images) in response to controlled images. Error bars represent the within-subject standard error of the mean, and stars indicate statistical significance.

parts of cortex had reliable preferences for each stimulus type, both at the group level (Figure 1b) and at the single-subject level (Supplementary Figure 3). Reachspace preferences (blue) were evident in three distinct zones: posterior ventral cortex, occipital-parietal cortex, and superior parietal cortex. These zones of preference lay adjacent to known object-preference zones (yellow), and scene-preference zones (green). Thus, while all three conditions extensively drive visual cortex, the activation landscapes differ in a systematic manner.

To estimate the magnitude of reachspace preferences, we defined reachspace-preferring regions of interest (ROIs) around the peaks in reachspace-preference appearing in anatomically consistent locations across subjects. Half of the data (activations from Image Set A) were submitted to a conjunction analysis to find voxels with a preference for reachspaces over objects, and reachspaces over scenes. This procedure yielded three reachspace-preferring ROIs: a ventral region (vROI), primarily located in the posterior collateral sulcus, an occipito-parietal region (opROI), variably located in the middle or superior occipital gyri, and a superior parietal region (spROI), in the anterior portion of the superior parietal lobe. Talairach (TAL) coordinates for these ROIs are given in Supplementary Table 1.

Next, we examined activation magnitude in the remaining half of the data (Image Set B), and found that reachspace views elicited significantly higher activations than both scenes and objects in all three ROIs (Figure 2, vROI: RS>O: $t(8)=5.33$, $p<0.001$, RS>S: $t(8)=4.66$, $p=0.001$; opROI: RS>O: $t(6)=5.20$, $p=0.001$, RS>S: $t(6)=4.55$, $p=0.002$; spROI: RS>O: $t(7)=6.16$, $p<0.001$, RS>S: $t(7)=5.22$, $p=0.001$). These results also held when swapping the image set used to define the ROIs and test for activation differences (see Supplementary Table 2 for all statistics). Finally, when the same data were broken out by semantic category, the reachspace categories elicited the highest overall responses (Figure 2).

Taken together, these analyses show that there are portions of cortex with systematically stronger responses to images of reachable-scale environments than to navigable-scale scenes and single object images.

Low-level Control and Replication. In Experiment 2, we aimed to replicate the finding that reachspaces elicit greater activity than scenes and objects in some regions, and to test whether the response preferences for reachspaces are attributable to factors beyond very simple feature differences. Twelve participants (2 of whom had completed Experiment 1) viewed Image Set A ("original" images), and a version of Image Set B that was matched in mean luminance, contrast, and global spatial frequency content ("controlled" images; see Figure 3a and Supplementary Figure 4 for examples).

Preference maps elicited by original and controlled images had highly similar spatial organization (Figure 3b; see Supplementary Figure 5 for single-subject maps). At the group level, 69.9% of visually-responsive voxels preferred the same condition, across original and controlled image formats (chance=33.3%, 50.3%+/-1.5% match at the single subject level; see Supplementary Figure 6). Further, the topographies found in Experiment 2 with original images also match those found in Experiment 1 (67.4% of voxels in group-level preference maps had the same preference).

Additionally, the ROI results replicated with controlled images. Specifically, ROIs were defined in Experiment 2 subjects using original images, and activations were extracted for controlled images (Figure 3c). Preferential responses to reachspaces were generally maintained (vROI: RS>O: $t(9)=2.08$, $p=0.034$, RS>S: $t(9)=2.72$, $p=0.012$; opROI: RS>O: $t(5)=2.38$, $p=0.032$; spROI: RS>O: $t(5)=3.61$, $p=0.008$, RS>S: $t(5)=2.02$, $p=0.05$; RS>S in opROI was not significant: $t(5)=0.79$, $p=0.234$). Note that in most of these ROIs, controlled images generally elicited lower overall activation magnitude than original images, and in some cases, the strength of the reachspace preference was slightly weaker than in the original image set (see Supplementary Table 3).

In summary, Experiment 2 found that the controlled image set elicited weaker but similar responses to object, reachspace, and scene images, indicating that these brain responses are not solely driven by stimulus differences in luminance, contrast, or global spatial frequency content.
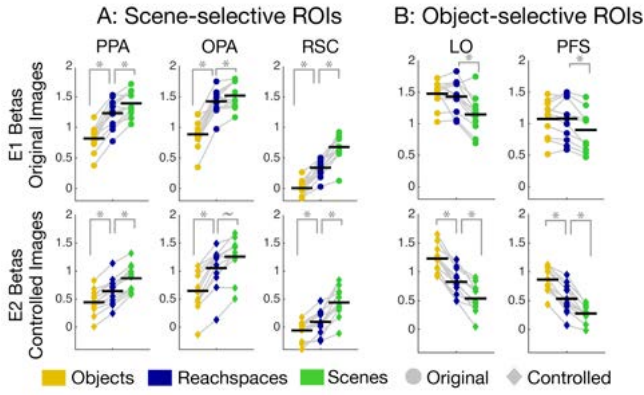
Fig. 4. Result for classic category-selective ROIs. A) Univariate response to objects, scenes, and reachspaces in scene-selective regions, for both original and controlled images (i.e. Experiment 1 and Experiment 2 respectively). B) Same analysis for object-selective regions.

Responses to reachable-scale environments in scene- and object-preferring regions. We next evaluated reachspace-evoked activity in scene- and object-selective regions using data from both Experiment 1 (original images) and Experiment 2 (controlled images). All category-selective ROIs were defined using independent localizer runs (see Supplementary Methods).

In scene-preferring regions (parahippocampal place area, PPA; occipital place area, OPA; retrosplenial cortex, RSC), reachspaces elicited an intermediate level of activation for both original and controlled images (Figure 4a). That is, reachspace images evoked stronger activation than object images (original images: PPA: t(11)=11.29, p<0.001; OPA: t(10)=9.16, p>0.001, RSC: t(11)=9.15, p<0.001; controlled images: PPA: t(11)=8.43, p<0.001; OPA: t(10)=9.32, p<0.001; RSC: t(11)=5.24, p<0.001), and weaker activation than scene images, although this difference was marginal in OPA for original images (original image set: PPA: t(11)=4.50, p<0.001, OPA: t(10)=1.63, p=0.067; RSC: t(11)=6.80, p<0.001; controlled images: PPA: t(11)=9.69, p<0.001; OPA: t(10)=4.25, p=0.001, RSC: t(11)=6.48, p<0.001; see Supplementary Table 2 for results in original images where the ROI-defining and activation-extracting runs were swapped; see Supplementary Table 3 for comparisons of activations evoked by original and controlled images).

In object-preferring regions (lateral occipital, LO; and posterior fusiform sulcus, pFs), reachspaces also showed intermediate activation levels of activation in most of the comparisons (Figure 4b). Specifically, reachspace images elicited significantly more activity than scene images (original images: LO: t(10)=5.55, p<0.001; pFs: t(10)=4.86, p<0.001; controlled images: LO: t(11)=8.10, p<0.001; pFs: t(11)=6.04, p<0.001). Additionally, reachspace images elicited significantly weaker activation than objects for controlled images (LO: t(11)=11.20, p<0.001; pFs: t(11)=12.19, p<0.001), but showed a similar overall activation with object images in their original format (LO: t(10)=0.86, p=0.204; pFs: t(10)=-0.12, p=0.547; see Supplementary Table 3 for all comparisons between activations to original vs. controlled images).

Taken together, these analyses show that reachspaces elicit an intermediate degree of activity in both scene- and object-preferring ROIs. These results provide further evidence that

views of near-scale environments evoke different cortical responses than both scene and objects images.

Functional signatures of reachspace-preferring cortex. Next, we examined how object-, scene-, and reachspace-preferring ROIs differ in their broader functional signatures. We first report two opportunistic analyses from Experiment 1, which leverage stimulus conditions present in our localizer runs, then, we report data from a new Experiment 3, with planned functional signature analyses.

In our first opportunistic analysis, we examined the responses of regions with object, reachspace, and scene preferences to the eccentricity conditions present in the Experiment 1 retinotopy protocol (Figure 5a; see also Supplementary Figure 7). Reachspace-preferring regions showed a peripheral bias, which was significant at a conservative post-hoc statistical level for the ventral and occipital reachspace regions, but not in the superior parietal region (Figure 5a; vROI: t(8)=3.90, p=0.005; opROI: t(6)=4.82, p=0.003; spROI: t(7)=3.29, p=0.013; two tailed post-hoc paired ttest with Bonferri-corrected alpha=0.006). Similarly, scene regions were strongly peripherally-biased (PPA: t(11)=17.59, p<0.001; OPA: t(10)=9.27, p<0.001; RSC: t(11)=12.49, p<0.001). In contrast, object regions showed mixed biases, which did not reach significance after Bonferroni correction (LO: foveal bias, t(10)=2.68, p=0.023; pFs: peripheral bias, t(10)=2.26, p=0.047). These results show that regions which responded preferentially to reachspaces, like scene-selective regions, are most sensitive to peripheral stimulation.

In our second opportunistic analysis, we investigated how ROIs differed in their response profile to a broad selection of categories present in the Experiment 1 localizer: faces, bodies, hands, objects, multiple objects, white noise, and scenes. Activations were extracted from reachspace-preferring ROIs. Since localizer runs were no longer available to define scene and object ROIs, these ROIs were approximated using a 9-mm radius sphere around their average TAL location, estimated based on a literature review. Activations for all regions are plotted as fingerprint profiles in Figure 5b.

In all three reachspace-preferring ROIs, images of multiple objects elicited the highest activation (significant in ventral and superior parietal ROIs: vROI: Multiple Objects>Scenes, t(8)=3.49, p<0.01; spROI: Multiple Objects>Bodies, t(7)=5.54, p<0.01; marginal in opROI: Multiple Objects>Scenes t(6)=2.32, p=0.03). In contrast, scene and object preferring ROIs showed different functional signatures. The approximated PPA and RSC regions preferred scenes over all other conditions, including Multiple Objects (Scenes>Multiple objects in PPA: t(11)=12.02, p> 0.001; in RS: t(11)=7.87, p> 0.001; one-tailed paired t-test, post hoc alpha level=0.02). This difference was not significant for approximated OPA (t(11)=-0.18, p=0.57). Finally, approximated LO and pFs regions showed maximal response to bodies, with broad tuning to hands, faces, object and multiple objects, and no differences between single objects and multiple objects in either ROI (LO: t(11)=-0.15, p=0.56; pFs: t(11)=-0.85, p=0.79). Overall, these exploratory analyses suggest that all three reachspace-preferring regions show a similar response profile with each other, in spite of their anatomical separation, and this profile is distinct that of from scene and object preferring regions.

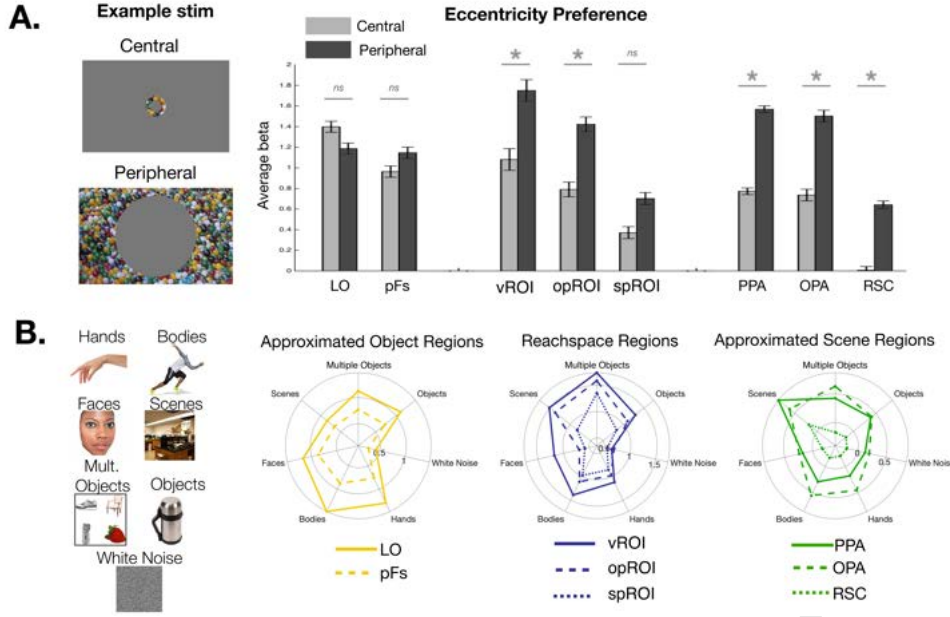To test this formally, Experiment 3 probed responses in

Fig. 5. Response properties of reachspace regions, compared with scene and object regions. A) Stimuli and results showing the eccentricity bias of object, reachspace, and scene-preferring areas. Error bars show the within-subjects standard error of the mean, and stars indicate statistical significance. B) Stimuli and results showing the profile of responses across a range of categories for reachspaces regions, and regions corresponding to the anatomical locations of object- and scene-selective areas. Beta values are plotted for each condition in a polar plot; negative values were thresholded to 0 for visibility.

²⁹¹ all ROIs to a broad range of conditions (Figure 6, stimuli in
²⁹² Supplementary Figure 8). These conditions included views
²⁹³ of standard reachspaces, objects and scenes, as well as four
²⁹⁴ different multi-object conditions (all depicting multiple ob-
²⁹⁵ jects with no background), and two different minimal object
²⁹⁶ conditions (depicting near-scale spatial layouts with one or no
²⁹⁷ objects). A final condition depicted vertical reachspaces, where
²⁹⁸ the disposition of objects was vertical rather than horizontal
²⁹⁹ (e.g. shelves, peg-boards). Experiment 3 was conducted in
³⁰⁰ the same session as Experiment 2, and involved the same par-
³⁰¹ ticipants and functionally-defined ROIs. Activations from all
³⁰² conditions were extracted from each ROI and the fingerprints
³⁰³ were compared.

³⁰⁴ Across these 10 conditions, reachspace-preferring regions
³⁰⁵ had a different fingerprint of activation than scene and object
³⁰⁶ regions (Figure 6a). To test the significance of the difference
³⁰⁷ in fingerprint profiles, responses across all conditions were
³⁰⁸ averaged over the reachspace ROIs to create a reachspace-ROI
³⁰⁹ fingerprint, then compared to the scene-ROI fingerprint (aver-
³¹⁰ aged over scene regions) and object-ROI fingerprint (averaged
³¹¹ over object regions) using a 2-way ANOVA. An omnibus test
³¹² of ROI-type (object, reachspace or scene) by condition re-
³¹³ vealed an ROI type-by-condition interaction (F(9, 329)=65.55,
³¹⁴ p<0.001), showing that the patterns of activations across the
³¹⁵ 10 conditions varied as a function of ROI- type. This dif-
³¹⁶ ference held when reachspace-ROIs were compared to scene-
³¹⁷ and object- ROIs separately (interaction effect for reachspace
³¹⁸ versus scene ROIs: F(9, 219)=32.20, p<0.001; for reachspace
³¹⁹ versus object ROIs: F(9, 219)=47.89, p<0.001). These results
³²⁰ further corroborate the conclusion that reachspace-preferring
³²¹ regions have a distinct representational signatures than object-
³²² and scene-preferring cortex.

³²³ Examining this response profile in more detail, in all three
³²⁴ reachspace-preferring ROIs, responses were higher to all multi-
³²⁵ object conditions (Figure 7, orange outline) than to empty
³²⁶ or single-object reachspaces (blue outline). To quantify this,
³²⁷ responses to all multi-object conditions were averaged, as were

³²⁸ responses to empty reachspaces and single-object reachspaces,
³²⁹ and two resulting activation levels were compared with a post
³³⁰ hoc t-test (vROI: t(9)=7.75, p <0.01; opROI: t(5)=4.57, p
³³¹ <0.01; spROI: t(5)=4.50, p < 0.01). This pattern of data sug-
³³² gests that the presence of multiple easily-individuated objects
³³³ may be particularly critical for driving the strong response to
³³⁴ reachspace images relative to full-scale scenes, where object
³³⁵ content may be less prominent than layout information. In
³³⁶ contrast, in scene-preferring regions, the empty reachspace
³³⁷ images generated higher responses than multiple object arrays,
³³⁸ although this difference was marginal in OPA (Supplemen-
³³⁹ tary Figure 9; PPA: t(11)=-8.16, p < 0.01; OPA: t(10)=-1.49,
³⁴⁰ p=0.08; RSC: t(11)=-7.28, p < 0.01). This result is consistent
³⁴¹ with prior work showing scene-regions strongly prefer empty
³⁴² full-scale rooms over multiple objects, and generally reflect
³⁴³ responses to spatial layout ((1)).

³⁴⁴ These activation profiles also illustrate how the stimuli
³⁴⁵ used to a define a region do not allow us to directly infer
³⁴⁶ what specific information is encoded there. For example, scene
³⁴⁷ images depict both spatial layout and multiple objects, but
³⁴⁸ scene ROIs are relatively more sensitive to the spatial layout
³⁴⁹ content of the images. Analogously, reachspace images depict
³⁵⁰ both spatial layout and multiple objects, but reachspaces ROIs
³⁵¹ are relatively more sensitive to the multi-object content of
³⁵² the images. Thus, the claim is not necessarily that these
³⁵³ are "reachspace-selective" regions. Rather, the claim is that
³⁵⁴ these regions are responsive to some content that is relatively
³⁵⁵ more present in naturalistic reachspace images than scene and
³⁵⁶ object images, and we suggest that the presence of multiple
³⁵⁷ individuated objects is likely to be an important factor. Future
³⁵⁸ work will be required to further articulate the distinctive roles
³⁵⁹ of these regions.

New territory vs. new subdivisions of scene-preferring regions. ³⁶⁰ Fi-
³⁶¹ nally, we conducted several targeted analyses aimed at un-
³⁶² derstanding whether reachspace-preferring regions are truly
³⁶³ separate regions of cortex from scene-and object-preferring
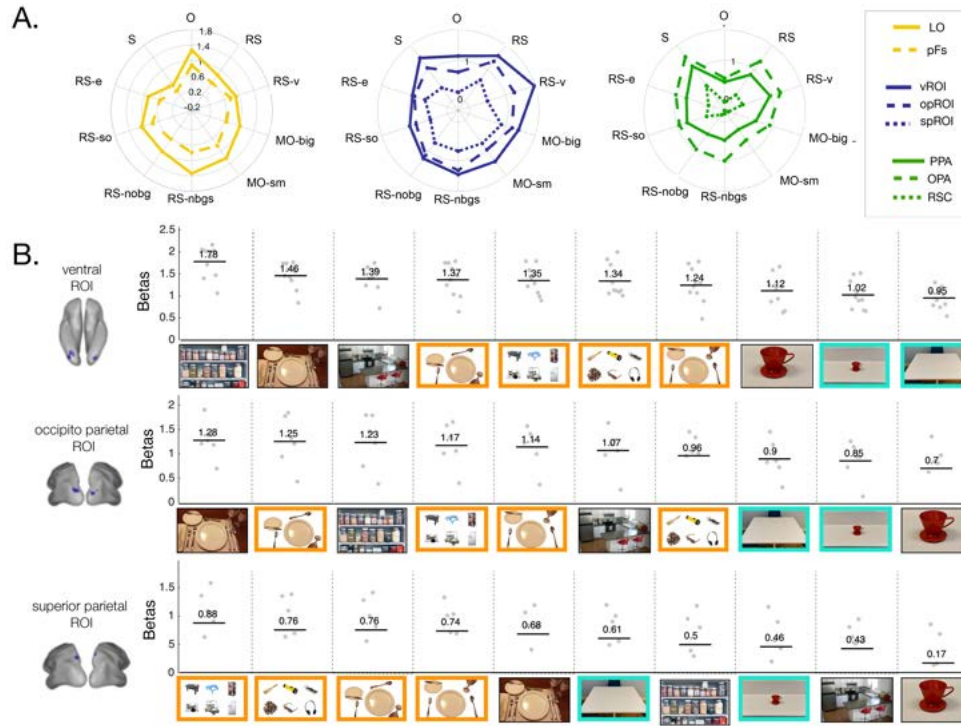³⁶⁴ ROIs, or whether they are simply new subdivisions. First,

Fig. 6. Experiment 3 results. A) Fingerprint profile of responses over all conditions in object, reachspace, and scene ROIs. O: singleton objects; RS: reachspaces; RS-v: vertical reachspaces; MO-b: multiple big objects; MO-sm: multiple small ojbects; RS-nbgs: reachspaces with no background, object positions scrambled; RS-nobg: Reachspace images with only objects with background removed; RS-so: reachspace images with only one object; RS-e: empty reachspace images. B) Responses in reachspace-preferring ROIs across all Experiment 3 conditions, plotted in order from highest to lowest activations. Images with orange borders indicate stimuli dominated by multiple objects, and images with teal borders highlight images of reachable space with low object content. See Supplement for all stimuli used in the experiment, in a larger format.

₃₆₅ we subdivided classically-localized PPA into anterior and pos-
₃₆₆ terior regions (24) (25), and found that neither subdivision
₃₆₇ showed a reachspace-preference (Supplementary Figure 10).
₃₆₈ These analyses indicate that the ventral reachspace-preferring
₃₆₉ ROI does not correspond to this known subdivision of PPA.

₃₇₀ Next, we quantified the overlap between all ROIs, given
₃₇₁ that it was statistically possible for scene-preferring regions
₃₇₂ (defined with a standard scene>object contrast in localizer
₃₇₃ runs) to overlap with reachspace-preferring regions (defined
₃₇₄ RS>O and RS>S conjunction contrast in experimental runs).
₃₇₅ However, we found relatively little overlap among the ROIs
₃₇₆ (e.g. for the ventral ROI, there was a 4.4 +/- 1.8% overlap
₃₇₇ with PPA, 4.6 +/- 2.1% with pFs, and 0.1 +/- 0.1% with
₃₇₈ FFA; see Supplementary Tables 4 and 5 for all overlap results).
₃₇₉ The relationship among these ROIs is visualized for three
₃₈₀ individual participants in Figure 7, and for all participants
₃₈₁ in Supplementary Figure 11. Overall, reachspace-preferring
₃₈₂ ROIs largely occupy different regions of cortex than object-,
₃₈₃ scene-, face-, and hand-selective cortex.

₃₈₄ Finally, we examined whether reachspace regions could be
₃₈₅ an artifact of population mixing. For example, it is possible
₃₈₆ that the ventral reachspace-preferring region actually reflects
₃₈₇ an intermixing of object-preferring neurons (similar to nearby
₃₈₈ pFs), and scene preferring neurons (similar to nearby PPA)
₃₈₉ whose competing responses to object and scene images av-
₃₉₀ erage out at the scale of fMRI, creating the appearance of
₃₉₁ reachspace tuning. If this were the case, then we would expect
₃₉₂ that the functional profile of the ventral reachspace region
₃₉₃ over the 10 Experiment 3 conditions could be predicted by
₃₉₄ a weighted combination of responses in scene- and object-
₃₉₅ preferring regions. However, this was not evident in the data
₃₉₆ (SI; Supplementary Figure 12): no mixture of pFS and PPA
₃₉₇ tuning could predict the preference for all four multi-object

₃₉₈ conditions over both single object conditions. Further, the
₃₉₉ superior parietal ROI is also informative, as this region shows
₄₀₀ both a reachspace preference and a functional fingerprint sim-
₄₀₁ ilar to other reachspace-preferring ROIs, but is anatomically
₄₀₂ far from object- or scene-selective regions.

₄₀₃ **Discussion.** The aim of this study was to characterize how
₄₀₄ the visual system responds to views of reachable environments
₄₀₅ relative to views of full-scale scenes and singleton objects.
₄₀₆ We found that (1), reachable environments activate distinct
₄₀₇ response topographies from both scenes and objects; (2) re-
₄₀₈ gions with reachspace-preferences are present in consistent
₄₀₉ locations across participants, allowing us to define ROIs in the
₄₁₀ posterior collateral sulcus, in dorsal occipto-parietal cortex,
₄₁₁ and the superior parietal lobule; (3) the response topogra-
₄₁₂ phies of reachspace preferences are maintained in an image
₄₁₃ set equating luminance, contrast, and global spatial frequency;
₄₁₄ (4) reachspaces elicit dissociable activity in scene and object
₄₁₅ ROIs, driving these regions to an intermediate degree; (5)
₄₁₆ reachspace-preferring regions have peripheral biases and (6)
₄₁₇ have distinctly higher response to the presence of multiple
₄₁₈ isolated objects over near-scale spatial layout with minimal
₄₁₉ object content, a combination that is unique among the ROIs
₄₂₀ explored here; (7) the reachspace-preferring regions do not
₄₂₁ appear to be a subdivision of classic category-selective areas.

₄₂₂ *Situating reachspace-preferring cortex.* Activations across a sim-
₄₂₃ ilar constellation of regions were found when participants
₄₂₄ attended to the reachability of objects versus their color or
₄₂₅ location (26), and when participants attended to a ball ap-
₄₂₆ proaching their hand versus a more distant target (27). In
₄₂₇ addition, the three reachspace-preferring ROIs appear to over-
₄₂₈ lap a subset parcels involved in making predictions about the
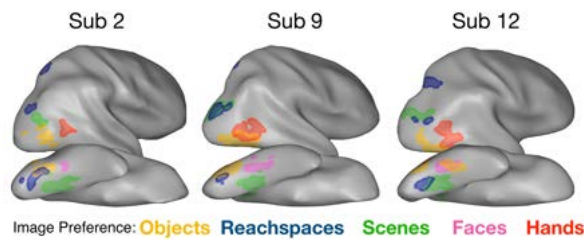₄₂₉ physical behavior of objects (28). Taken together, the corre-

Fig. 7. Depiction of the location of the reachspace ROIs in relation to scene-, object-, and face-preferring ROIs, shown in the right hemispheres of three example subjects.

spondence between these results suggest that the ROIs which preferred reachable-scale views may be generally important for reachability judgments, and suggest a potentially broader role in behaviors that rely on accurate predictions regarding objects in the physical world.

The ventral reachspace ROI lies near a swath of cortex sensitive to features of object ensembles (29), to the texture and surface properties of single objects (30), to regions that are sensitive to videos of actions being performed in the near space (31), and near the posterior edge of a color-biased band running along ventral IT cortex (32) (23). The occipital reachspace ROI lies in the vicinity of inferior parietal regions associated with the maintenance of multiple objects in working memory (33). And, the superior parietal reachspace ROI falls near territory thought to contain information about the reachability of an object (34), and the type of object-directed hand movement that is planned (35). Interestingly, this ROI also appears to overlap the posterior locus of the Multiple-Demand Network, a network of fronto-parietal regions associated with the control of visual attention and the sequencing of cognitive operations (36) (37). Future studies with targeted comparisons will be required to map these functions together and assess the degree to which they draw on common representations.

Finally, it was recently found (31) that tuning of ventral and dorsal stream responses to videos of people performing actions was related to the "interaction envelope" (38) of the depicted action, and were sensitive to whether the actions were directed at objects in near space or far space. This result is also broadly consistent with the present results, where the scale of depicted space seems to be an important factor in the structure of responses across the entire visual system.

Implications for the visual representation of reachable space. The existence of reachspace-preferring cortex suggests that near-scale environments require some distinctive processing relative to navigable-scale scenes and close-scale objects. Part of these differences may relate to differences in scale between the views: perceived depth has been shown to affect activation strength in scene regions (39) (40) (41). However, it is clear that the ROIs which prefer reachspaces do not do so on the basis of scale alone: environments which were near-scale, but contained one or no objects elicited low responses in these regions. Instead, the regions responded strongly to images of multi-objects arrays, suggesting a role for object-related content. Is it possible, then, that these regions are best characterized as "multiple object regions"? How important are the background spatial components, such as the desktops, and texture cues to the perceived depth of the scene? Future work will be needed to characterize the effects of scale, number of objects, and their

interactions to clarify in these regions.

Finally, it is possible to extend theoretical frameworks for the large-scale organization of (isolated) object information, and apply them to the large-scale organization of object, reachspace, and scene views. For example, some have argued that the visual world is divided into domains linked to behavioral relevance, which are separately arrayed along the cortical sheet(20) (23). Consistent with this action-based perspective, objects, reachspaces and scenes differ in the kinds of high-level goals and behaviors they afford: objects afford grasping, reachspaces afford the coordinated use of multiple objects, and scenes afford locomotion and navigation. Others have argued that the large-scale organization is more of an emergent property that follows from experienced eccentricity and aggregated differences in mid-level image statistics (21) (22) (42) (23). Consistent with this input-based perspective, reachspace images as a class are perceptually distinct from both scene and object images, a distinction which is also evident in the learned representations of deep neural networks (18). In sum, there are both action-based and image feature properties that can jointly motivate a large-scale division of objects, reachspaces, and scenes across the visual system.

Materials and Methods

In-text methods provide details about subject, stimuli and ROI definitions. All other method details are available in the supplement.

Subjects. Twelve participants were recruited for Experiment 1 and Experiment 2. Two participants completed both. Experiment 3 was conducted in the same session as Experiment 2. All participants gave informed consent and were compensated for their participation. All procedures were approved by the Harvard University Human Subjects Institutional Review Board.

Stimuli. All stimuli are available on the Open Science Framework (https://osf.io/g9aj5/). For Experiment 1, we collected views of objects, scenes, and reachable environments, each with 10 images from 6 semantic categories (bar, bathroom, dining room, kitchen, office, art studio), yielding 60 images per scale. These images were divided into 2 equal sets—Image Set A and Image Set B. Object images depicted close-scale views (within 8-12 inches from the object) on their natural background, e.g.: a view of a sponge with a small amount of granite countertop visible beyond it. Reachspace images depicted near-scale environments that were approximately as deep as arm's reach (3-4ft), and consisted of multiple small objects arrayed on a horizontal surface, e.g.: a knife, cutting board and an onion arrayed on kitchen counter. Scene images depicted views of the interior of rooms, e.g.: a view of a home office. For Experiment 2, we created a controlled version of Image Set B where all images were grayscaled, matched in average luminance, contrast, and global spatial frequency content using the SHINE toolbox ((43)). Experiment 3 included 10 stimulus conditions: (1) reachspaces images with the background removed in photoshop, yielding images of multiple objects in realistic spatial arrangements; (2) reachspaces images with background removed and the remaining objects scrambled, where the objects from the previous condition were moved around the image to disrupt the realistic spatial arrangement; (3) 6 objects with large real-world size, e.g. trampoline, dresser, arranged in a 3x2 grid on a white background; (4) 6 objects with small real world size, e.g. mug, watch, arranged in a 3x2 grid on a white background, and presented at the same visual size as the previous image condition; (5) reachable environments with all objects removed except the support surface; (6) reachspaces containing only a single object on the support surface; (7) vertical reachspaces, where the disposition of objects was vertical rather than horizontal, e.g. shelves, peg-boards; (8) regular reachspaces, i.e. horizontal, as in earlier experiments (9) objects, i.e. close-up views of single objects on their natural background; and (10) scenes, i.e. navigable scale

environments. Further details on stimulus selection and controls are available in the supplement.

Defining ROIs with reachspace preferences.. For Experiment 1, Reachspace-preferring ROIs were defined manually in Brain Voyager by applying the conjunction contrast RS>O and RS>S, using four experimental runs with the same image set. We decided a priori to define all reachspace ROIs using Image Set A runs, and extract all activations for further analysis from Image Set B runs. These results are reported in the paper, but we also validated all analyses by reversing which image set was used to localize vs extract activations, and these results are reported in the supplement. For the ROIs used in Experiments 2 and 3 (run in the same session), we designed an automatic ROI-selection algorithm, guided by the anatomical locations of these regions in Experiment 1. This method allowed for the precise location of ROIs to vary over individuals, while still requiring them to fall within anatomically-constrained zones. The algorithm located the largest patch in the vicinity of the average location of the ROIs from E1 where the univariate preference for reachspaces over the next-most-preferred category exceeded 0.2 beta (more details in supplement). This automated procedure was developed using a separate pilot data set and all parameters were decided a priori (but see Supplementary Figure 8 for a visualization of the consequences of this parameter choice).

1. R Epstein, N Kanwisher, A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
2. DJ Kravitz, CS Peng, CI Baker, Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J. Neurosci.* **31**, 7322–7333 (2011).
3. MD Lescroart, JL Gallant, Human scene-selective areas represent 3d configurations of surfaces. *Neuron* **101**, 178–192 (2019).
4. S Park, TF Brady, MR Greene, A Oliva, Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J. Neurosci.* **31**, 1333–1340 (2011).
5. MF Bonner, RA Epstein, Coding of navigational affordances in the human visual system. *Proc. Natl. Acad. Sci.* **114**, 4793–4798 (2017).
6. FS Kamps, JB Julian, J Kubilius, N Kanwisher, DD Dilks, The occipital place area represents the local elements of scenes. *Neuroimage* **132**, 417–424 (2016).
7. SA Marchette, LK Vass, J Ryan, RA Epstein, Outside looking in: landmark generalization in the human navigational system. *J. Neurosci.* **35**, 14896–14908 (2015).
8. LK Vass, RA Epstein, Abstract representations of location and facing direction in the human brain. *J. Neurosci.* **33**, 6133–6142 (2013).
9. K Grill-Spector, Z Kourtzi, N Kanwisher, The lateral occipital complex and its role in object recognition. *Vis. research* **41**, 1409–1422 (2001).
10. K Grill-Spector, T Kushnir, S Edelman, Y Itzchak, R Malach, Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* **21**, 191–202 (1998).
11. FH Previc, The neuropsychology of 3-d space. *Psychol. bulletin* **124**, 123 (1998).
12. OJ Grüsser, Multimodal structure of the extrapersonal space in *Spatially oriented behavior*. (Springer), pp. 327–352 (1983).
13. G Rizzolatti, R Camarda, Neural circuits for spatial attention and unilateral neglect in *Advances in psychology*. (Elsevier) Vol. 45, pp. 289–313 (1987).
14. G di Pellegrino, E Làdavas, Peripersonal space in the brain. *Neuropsychologia* **66**, 126–133 (2015).
15. MS Graziano, CG Gross, The representation of extrapersonal space: A possible role for bimodal, visual-tactile neurons. *The cognitive neurosciences* **1**, 1021–1034 (1995).
16. E Maguire, The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand. journal psychology* **42**, 225–238 (2001).
17. A Torralba, A Oliva, Depth estimation from image structure. *IEEE Transactions on pattern analysis machine intelligence* **24**, 1226–1238 (2002).
18. EL Josephs, T Konkle, Perceptual dissociations among views of objects, scenes, and reachable spaces. *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 715 (2019).
19. HPO de Beeck, I Pillet, JB Ritchie, Factors determining where category-selective areas emerge in visual cortex. *Trends cognitive sciences* **23**, 784–797 (2019).
20. BZ Mahon, A Caramazza, What drives the organization of object knowledge in the brain? *Trends cognitive sciences* **15**, 97–103 (2011).
21. T Konkle, A Oliva, A real-world size organization of object responses in occipitotemporal cortex. *Neuron* **74**, 1114–1124 (2012).
22. MJ Arcaro, MS Livingstone, A hierarchical, retinotopic proto-organization of the primate visual system at birth. *Elife* **6**, e26196 (2017).
23. BR Conway, The organization and operation of inferior temporal cortex. *Annu. Rev. Vis. Sci.* **4**, 381–402 (2018).
24. M Bar, E Aminoff, Cortical analysis of visual context. *Neuron* **38**, 347–358 (2003).
25. C Baldassano, DM Beck, L Fei-Fei, Differential connectivity within the parahippocampal place area. *Neuroimage* **75**, 228–237 (2013).
26. A Bartolo, et al., Contribution of the motor system to the perception of reachable space: an fmri study. *Eur. J. Neurosci.* **40**, 3807–3817 (2014).
27. TR Makin, NP Holmes, E Zohary, Is that near my hand? multisensory representation of peripersonal space in human intraparietal sulcus. *J. Neurosci.* **27**, 731–740 (2007).
28. J Fischer, JG Mikhael, JB Tenenbaum, N Kanwisher, Functional neuroanatomy of intuitive physical inference. *Proc. national academy sciences* **113**, E5072–E5081 (2016).
29. JS Cant, Y Xu, Object ensemble processing in human anterior-medial ventral visual cortex. *J. Neurosci.* **32**, 7685–7700 (2012).
30. JS Cant, SR Arnott, MA Goodale, fmr-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Exp. Brain Res.* **192**, 391–405 (2009).
31. L Tarhan, T Konkle, Sociality and interaction envelope organize visual action representations. *Nat. Commun.* **11**, 1–11 (2020).
32. R Lafer-Sousa, BR Conway, NG Kanwisher, Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *The J. Neurosci.* **36**, 1682–1697 (2016).
33. Y Xu, Distinctive neural mechanisms supporting visual object individuation and identification. *J. Cogn. Neurosci.* **21**, 511–518 (2009).
34. JP Gallivan, C Cavina-Pratesi, JC Culham, Is that within reach? fmri reveals that the human superior parieto-occipital cortex encodes objects reachable by the hand. *J. Neurosci.* **29**, 4381–4391 (2009).
35. JP Gallivan, DA McLean, KF Valyear, CE Pettypiece, JC Culham, Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *J. Neurosci.* **31**, 9599–9610 (2011).
36. J Duncan, The multiple-demand (md) system of the primate brain: mental programs for intelligent behaviour. *Trends cognitive sciences* **14**, 172–179 (2010).
37. E Fedorenko, J Duncan, N Kanwisher, Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci.* **110**, 16616–16621 (2013).
38. WA Bainbridge, A Oliva, Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage* **122**, 408–416 (2015).
39. AS Persichetti, DD Dilks, Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex* **77**, 155–163 (2016).
40. MD Lescroart, DE Stansbury, JL Gallant, Fourier power, subjective distance, and object categories all provide plausible models of bold responses in scene-selective visual areas. *Front. computational neuroscience* **9**, 135 (2015).
41. JM Henderson, CL Larson, DC Zhu, Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: an fmri study. *Brain cognition* **66**, 40–49 (2008).
42. B Long, CP Yu, T Konkle, Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc. Natl. Acad. Sci.* **115**, E9015–E9024 (2018).
43. V Willenbockel, et al., The shine toolbox for controlling low-level image properties. *hist* **1**, 2 (2010).

Supplementary Information

Large-scale dissociations between views of objects, scenes, and reachable-scale environments in visual cortex

Emilie L. Josephs & Talia Konkle

Supplementary Methods

Subjects. Twelve participants were recruited each for Experiment 1 and Experiment 2. Experiment 3 was conducted in the same session as Experiment 2, and thus represents the same participants. Two people participated in both E1 and E2, and author EJ participated in E1. Participants were between the ages of 20 and 31, and 13 out of 22 participants were female. One additional person participated in E1, but was excluded prior to analysis for falling asleep in the scanner. All participants gave informed consent and were compensated for their participation. All procedures were approved by the Harvard University Human Subjects Institutional Review Board.

Acquisition and Pre-processing. All neuroimaging data were collected at the Harvard Center for Brain Sciences using a 32-channel phased-array head coil with a 3T Siemens Magnetom Prisma fMRI Scanner (Siemens Healthcare, Erlangen, Germany). The Siemens Auto-Align tool was used to ensure reproducible placement of image fields of view. High-resolution anatomical images were collected with a T1-weighted magnetization-prepared rapid gradient multi-echo sequence (multi-echo MPRAGE [1], 176 sagittal slices, TR=2530 ms, TEs=1.69, 3.55, 5.41, and 7.27 ms, TI=1100 ms, flip angle=7°, 1 mm3 voxels, FOV=256 mm, GRAPPA acceleration=2). For functional runs, blood oxygenation level-dependent (BOLD) data were collected via a T2*-weighted echo-planar imaging (EPI) pulse sequence that employed multiband RF pulses and Simultaneous Multi-Slice (SMS) acquisition [2-5]. For the task runs, the EPI parameters were: 69 interleaved axial-oblique slices (25 degrees toward coronal from ACPC alignment), TR=1500 ms, TE=30 ms, flip angle=75°, 2.0 mm isotropic voxels, FOV=208 mm, in-plane acceleration factor (GRAPPA)=2, SMS factor=3). The SMS-EPI acquisition used the CMRR-MB pulse sequence from the University of Minnesota.

Functional data were preprocessed using Brain Voyager QX software with MATLAB scripting. Preprocessing included slice-time correction (ascending trilinear interpolation), 3D motion correction (sinc interpolation), linear trend removal, temporal high-pass filtering (0.0078 Hz cutoff), spatial smoothing (4 mm FWHM kernel), AC-PC alignment and transformation into Talairach (TAL) coordinates. Three dimensional models of each subject's cortical surface were generated from the high-resolution T1-weighted anatomical scan using the default segmentation procedures in FreeSurfer. For visualizing activations on inflated brains, surfaces were imported into Brain Voyager and inflated using the BV surface module. Gray matter masks were defined in the volume based on the Freesurfer cortex segmentations.

E1 and E2 Stimuli. All stimuli are available for download on the Open Science Framework (https://osf.io/g9aj5/). We collected views of objects, scenes, and reachable environments, each with 10 images from 6 semantic categories (bar, bathroom, dining room, kitchen, office, art studio), yielding 60 images per scale. These images were divided into equal 2 sets—Image Set A and Image Set B. Object images depicted close-scale views (within 8-12 inches from the object) on their natural background, e.g.: a view of a sponge with a small amount of granite countertop visible beyond it. Reachspace images depicted near-scale environments that were approximately as deep as arm's reach (3-4ft), and consisted of multiple

small objects arrayed on a horizontal surface, e.g.: a knife, cutting board and onion arrayed on kitchen counter. Scene images depicted views of the interior of rooms, e.g.: a view of a home office.

Additionally, using Image Set B we created a controlled image set, where all images were grayscaled, matched in average luminance, contrast, and global spatial frequency content using the SHINE toolbox (Willenbockel, et al, 2010). Images were spatial frequency-matched using the specMatch function, then luminance-matched using the histMatch function, both with default parameters.

Experiment 1 reachspace preference analyses. The main experimental protocol for Experiment 1 consisted of a blocked design with 18 image conditions, depicting three scales of space (object, reachspace, scene views), drawn from six different semantic categories (bar, bathroom, dining room, kitchen, office, studio). Each run contained two blocks per condition, with blocks lasting 6s and consisting of 5 unique images and 1 repeated image. Within a block, each image was presented in isolation on a uniform gray background for 800ms followed by a 200ms blank. There were twelve 10s fixation blocks interleaved throughout the experiment, and each run started and ended with a 10s fixation block. A single run lasted 5.93 min (178 volumes). Participants viewed eight runs of the experimental protocol. Four runs were completed with Image Set A and four with Image Set B (see Stimuli), presented in alternating order over the course of the scan session. Participants' task was to detect a image repeated back-to-back, which happened once per block.

General linear models (GLMs) were computed using Brain Voyager software. In Experiment 1, for each participant, separate GLMs were fit for runs containing Image Set A and Image Set B, and a third GLM was fit to all experimental runs together (this combined GLM was only used for Experiment 1 preference map analysis). Data were modeled first with 3 condition regressors (object, reachspace, scene), and then again with 18 condition regressors (3 scales of space x 6 semantic category) for the finer-grained analyses by category and the searchlight analysis. The regressors were constructed based on boxcar functions for each condition, convolved with a canonical hemodynamic response function, and were used to fit voxel-wise time course data with percent signal change normalization and correction for serial correlations. The beta parameter estimates from the GLM were used as measures of activation to each condition for all subsequent analyses.

Preference Mapping. Group-level preference maps were computed by extracting responses to objects, reachspaces and scenes in each voxel from single-subject GLMs, then averaging over subjects. The preferred condition for each voxel was identified in the group average, and the degree of preference was computed as the activation differences (in betas) between the most preferred condition and the next-most-preferred condition. Responses were visualized for visually-responsive voxels only, which were defined as those that were active in an All vs Rest contrast at a threshold of t>2.0 in at least 30% of the participants. Early visual regions (V1-V3) were defined by hand on inflated brain, guided by the contrast of horizontal vs. vertical meridians from a retinotopy run (see below for run details). Group average V1-V3 was obtained by generating single-subject early visual cortex maps, and selecting voxels that fell within V1-V3 in at least 30% of the participants. These voxels were removed from the visualization. Preference maps were visualized by projecting these voxels' preferred condition (indicated by color hue) and the degree of preference (indicated by color intensity) onto the cortical surface of a sample participant. For Experiment 1, preference maps were computed from a GLM modeled with data from all 8 experimental runs.

RS ROI definition: For Experiment 1, three reachspace-preferring ROIs were defined manually in Brain Voyager by applying the conjunction contrast RS>O & RS>S, using four experimental runs with the same image set. Conjunction contrasts reveal voxels that show both a preference for reachspaces over scenes and reachspace over objects (assigning them the statistical value corresponding to the less robust of those contrasts). We had decided a priori to define all reachspace ROIs using Image Set A runs, and extract all activations for further analysis from Image Set B runs. These results are reported in the paper, but we

also validated all analyses by reversing which image set was used to localize vs extract activations, and these results are reported in this supplement.

Region-of-Interest Analysis. For ROI-based analyses, univariate activations were obtained by taking the average beta for each condition in each ROI, then averaging over subjects to obtain the group-level activations. Reachspace-preferring ROIs were defined from 4 runs of the experimental protocol, and activation were extracted from the remaining 4 runs. Experiment 1 activations were examined at two levels, with separate GLMs generated for each. At the environment-scale level we examined the activations to 3 conditions: objects, reachspaces and scenes. In each ROI, we tested whether the preferred condition activated the ROI significantly more than the other conditions, using a priori paired one-sided t-tests. We also extracted responses to objects, scenes and reachspaces at the more granular scale-by-category level (18 conditions: 6 semantic categories represented at each of 3 scales). These data were visualized in a bar graph, where the bars are ordered by the strength of the activation.

Experiment 2 reachspace preference analyses. The main experimental protocol for Experiment 2 was the same as Experiment 1. Four runs used original images, specifically Image Set A from Experiment 1, and four runs used controlled images, specifically Image Set B with the low level controls described above. Experiment 2 GLMs were computed as above, and the data were modeled with 3 condition regressors (object, reachspace, scene).

Preference mapping. Experiment 2 preference mapping used the same procedure as Experiment 1, with the difference that V1-V3 were not removed from the visualization and subsequent quantification, since these regions were not localized in E2. For Experiment 2, separate preference maps were computed from original- and controlled-image runs, each estimate from a GLM with 4 runs.

To quantitively compare the similarity of preference maps elicited by original and controlled images, we assessed the proportion of voxels that showed the same preference across image sets. For the group-level preference map, we first extracted the preferred category and the strength of that preference for each voxel within the group-level mask, for original and controlled images separately (as described above). Next, we extracted the number of voxels that showed the same preference in the two maps (original vs controlled), then divided this by the total number of voxels in the visually-evoked mask, to obtain the proportion of voxels with consistent preference over the image sets. For group-level comparisons, we performed the analyses above on single-subject data, then averaged these values over all subjects.

This method was additionally used to compute the replicability of the original image activations between Experiment 1 and Experiment 2. To do so, the preference map for Experiment 1 was generated from Image Set B only, so that the preference maps being compared were generated from GLM parameter estimates made using the same amount of data (4 runs). Additionally, a common activity mask for the two preference maps was defined by taking the voxels that showed an All-vs-Rest preference of t>2.0 betas in 60% of all of the subject included (i.e. E1 and E2 subjects). Since this analysis was between subjects and no within-subject comparisons were available, the match in preference maps across experiments was only computed at the group-level.

RS ROI definition: For Experiment 2, we designed an automatic ROI-selection algorithm, guided by the locations of these regions in Experiment 1. ROIs were defined separately for each participant using the following procedure. First, a spherical proto-ROI was defined around the average central locations of each ROI from E1. The size of the proto-ROI was set to a radius of 6 voxels (18 mm) for the ventral and superior parietal patches, and 9 voxels (27mm) for the occipital patch, to account for different amounts of variation in the expected ROI locations. Then, the reachspace conjunction map with RS>O & RS>S was computed and spatially smoothed (5-mm gaussian kernel, sigma=1). Next, the single voxel with the highest reachspace-preference falling in each proto-ROI was selected and used as the center of 6mm spherical ROI.

Finally, the voxels within this sphere with the most statistically robust preference for reachspaces were retained for the final ROI, using the following procedure. Low-preference voxels were iteratively dropped from the ROI until the region's univariate preference for reachspaces over the next-most-preferred category exceeded 0.2 beta. This method allowed us to define the largest ROI that still showed a relatively high reachspace bias. This automatic ROI-selection regime was developed in a separate pilot data set before being applied here, and all parameters were decided a priori, but see **Supplementary Figure 13** for visualization of how parameter variation affected significance of the final analysis. Activations within these ROIs were always assessed from independent data sets.

Region-of-Interest Analysis: ROI analysis of reachspace regions used the same procedure as Experiment 1, with the exception that this analysis was only performed at the environment scale level (i.e. betas were extracted for objects, reachspaces, and scenes separately, pooling over semantic category). ROIs were defined with the 4 runs depicting original images, and activations were extracted from the 4 controlled-image runs, and compared using a priori paired one-sided t-test.

Experiment 1 classic category-selective ROI analysis. Classic category-selective ROIs were defined in Experiment 1 using a standard localizer protocol. Stimuli included images of bodies, faces, hands, objects, multiple objects, scenes, and white noise. Body images showed clothed bodies with the head erased in photoshop, in a variety of poses. Face images were cropped from the chin to the top of the head, and depicted a variety of facial expressions from humans of different ages, races, and genders. Object images showed single objects on a white background. Multi-object images showed four randomly-selected unique objects occupying four quadrants around the center fixation location, presented over a white background. Scene images showed indoor and outdoor images of navigable-scale spaces.

The localizer protocol contained 8 blocks per image condition, with blocks lasting 6s and consisting of 5 unique images and 1 repeated image. All images were presented in isolation on a uniform gray background for 800ms followed by a 200ms blank. There were eight 8s fixation blocks interleaved throughout the experiment, and each run started and ended with an additional 8s fixation block. A single run lasted 6.9 min (208 volumes), and participants viewed four runs of the localizer protocol. Participants' task was to detect an image repeated back-to-back, which happened once per block.

ROIs were defined using standard contrasts, and ROI activations were extracted from 4 runs of the main experimental protocol using the same univariate approach described above. Activations were extracted separately for Image Set A runs and Image Set B runs. The latter are reported in the paper, as explained above, and the former appear in the supplement. All stats were a priori paired one-sided t-test.

Reachspace ROI overlap analysis. In order to quantify whether the reachspace ROIs consistently overlapped any of the classic ROIs, we first divided them into ventral (PPA, pFs, vRS), and lateral-dorsal ROIs (OPA, LO, hand-preferring, opRS, spRS). Next, we assessed the overlap between the RS ROIs in a given division with each of the other ROIs in that division. For each subject, whole-brain masks were created for each ROI in the pair under comparison, and the number of voxels appearing in both masks was extracted. Then, the number of overlapping voxels was divided by the total number of voxels in the reachspace region, to obtain the percentage of the reachspace ROI voxels that overlapped the comparison ROI. With this definition, overlap estimates of 100% indicate that the reachspace-preferring regions fall fully into existing known regions; estimates of 0% indicate complete separation. This was computed separately for each hemisphere, and for RS ROIs created from each image set (Image Set A vs Image Set B).

Experiment 2 classic ROI analysis. Scene- and object-selective regions were defined in Experiment 2 from the main experimental protocol runs with the original images: LO and pFs were defined as objects>scenes; PPA, OPA and RSC were defined as scenes>objects. Activations from all regions were

extracted from the 4 experimental runs depicting controlled images, and the analysis was otherwise carried out as described in Experiment 1.

Original vs Controlled comparisons. Differences in the patterns obtained in ROI responses between original images (Experiment 1) and controlled images (Experiment 2) were assessed in two ways. First, the overall difference in activations between the images sets was assessed by averaging all the activations (object, reachspace, and scenes) within an ROI to obtain its mean response. This was compared between the experiments using between-subject ttest. Statistical threshold were set using Bonferroni corrections, where the number of comparisons was taken as the number of ROIs of each type (i.e. reachspace preferring, object-preferring, and scene-preferring). Second, we examined whether the magnitude of the differences between conditions was different for the two image sets. For this, we calculated the difference between reachspaces and scenes for a given ROI across all subjects, then averaged over subjects. The size of this difference was then compared between original and controlled images. The same was then performed for objects. Comparisons used between-subject t-tests, and statistical threshold were set using Bonferroni corrections, where the number of comparisons was taken as the number of ROIs of each type multiplied by 2 (since there are two comparisons per ROI: RS vs S and RS vs O).

Experiment 1 and Experiment 2 PPA subdivision. For Experiment 1 and 2, we additionally subdivided PPA. For each subject, we separately split the PPA from the left and right hemisphere at the midpoint along the anterior to posterior axis. Anterior and posterior PPA were then submitted to the same ROI analysis described above. Statistical tests were completed using Bonferroni-corrected paired one-sided t-test with alpha 0.0125 (i.e. 0.05/4, reflecting the four comparisons being performed in each image set).

Eccentricity profile analysis. Data for the eccentricity analysis were collected in the same run as Experiment 1, and thus represent the same subjects and reachspace ROIs. The retinotopy protocol consisted of 4 stimulus conditions: horizontal bands, vertical bands, central stimulation, and peripheral stimulation. Vertical and horizontal bands (subtending ~$1.7° × 22°$ and ~$22° × 1.7°$ respectively) showed checkerboards which cycled between states of black-and white, white-and-black, and randomly colored at 6hz. Central and peripheral rings (radius ~$1.2°$ to $2.4°$ and radius ~$9.3°$ to the edges of the screen, respectively) showed photograph fragments which cycled between patterns of object ensembles (e.g. beads, berries, buttons) and scene fragments (c.f. Cant & Xu, 2012; Zeidman, Silson, Schwarzkopf, Baker & Penny, 2018). Each run contained 5 blocks per condition, with blocks lasting 12 seconds. There were four 12s fixation blocks interleaved throughout the experiment, and each run started and ended with an additional 12s fixation block. Each run lasted 4.4 min (162 volumes), and participants viewed two runs of the retinotopy protocol. Participants' task was to maintain fixation and press a button when the fixation dot turned red, which happened at a random time once per block.

ROI analysis. We explored the eccentricity preference of object, reachspace, and scene ROIs (defined as described for Experiment 1 above), and for ROIs corresponding to scenes- and object ROIs. Average betas were extracted for two eccentricity conditions: central stimulation, and peripheral stimulation. Activations in the two conditions were compared using a paired one sided t-test with a Bonferroni corrected alpha value of 0.0125 (i.e. 0.05/8, reflecting the 8 ROIs where we tested for a difference between these conditions).

Post-hoc fingerprint profile analysis. To examine the broader tuning of object, scene and reachspace ROIs, we performed a post-hoc analysis, using activations extracted from the Experiment 1 localizer. The localizer runs included bodies, faces, hands, objects, multiple objects, scenes, and white noise). We extracted responses in Experiment 1 reachspace-preferring ROIs to these 8 conditions for each subject, and averaged the activations over subjects. First, we visualized these responses in a polar plot. Next, we noted what the most preferred condition was, and tested whether this was significantly different

than the next-most preferred condition using one-tailed pair t-tests with Bonferroni-corrected alpha 0.0167 (0.05/3, reflecting the 3 reachspace-preferring ROIs where we tested for this difference).

We also examined activations to these conditions in object- and scene-processing cortex. Since the localizer runs were used to extract activations, we couldn't use them to define ROIs. Instead, ROIs were defined as a spherical ROI with a 9-mm radius centered on the typical anatomical location of each region based on an internal meta-analysis (for left/right hemisphere, ROI centers were as follows: PPA: -25 -41 -6/ 25 -42 -7; OPA: -25 -76 25/ 28 -81 26; RSC: -16 -51 9/ 18 -49 8; LO: -39 -71 -4/ 41 -68 -4; pFs: -38 -53 -13/ 38 -50 -14).The difference between the preferred and next-most-preferred condition was assess using one-tailed pair t-tests with Bonferroni-corrected alpha 0.025, and 0.0167 respectively for areas corresponding to the anatomical location of object and scenes ROIs.

Experiment 3 fingerprint profile analysis. Experiment 3 stimuli contained 10 conditions intended to further probe the response profile of the reachspace regions. The conditions were the following: 1) reachspaces images with the background removed in photoshop, yielding images of multiple objects in realistic spatial arrangements; 2) reachspaces images with background removed and the remaining objects scrambled, where the objects from the previous condition were moved around the image to disrupt the realistic spatial arrangement; 3) 6 objects with large real-world size (e.g. trampoline, dresser) arranged in a 3x2 grid one a white background; 4) 6 objects with small real world size (e.g. mug, watch) arranged in a 3x2 grid a white background (presented at the same visual size as the previous image condition); 5) reachable environments with all objects removed except the support surface; 6) reachspaces containing only a single object on the support surface; 7) vertical reachspaces, where the disposition of objects was vertical rather than horizontal (e.g. shelves, peg-boards); 8) regular (i.e. horizontal) reachspaces; 9) objects (i.e. close-up views of single objects on their natural background); and 10) scenes (i.e. navigable scale environments).

Images from conditions 1 and 2 above (reachspace with background removed, and reachspace with background removed and remaining objects scrambled) were generated from the same original images. First, condition 1 images were generated by selecting high-quality reachspace images, and erasing all image content except the 6 salient objects which conveyed the identity and layout of the space. Then, condition 2 images were generated by scrambling the arrangement of the 6 remining objects in the image, and occasionally rotating objects, to break the sense of spatial congruity among them. We ensured that the average placement of objects across all the images (i.e. the heatmap of object locations) was equivalent between condition 1 and condition 2. Images in conditions 5, 6 and 9 (empty reachspaces, reachspaces with single objects, and close up view of single objects) were taken by the experimenter, and represented the same environments. Specifically, a suitable reachspace was selected by the experimenter and cleared of all objects for condition 5, and an images was taken with a camera on a tripod. Then a single salient object was placed in the center of the reachspace for condition 6, at which point a second picture was taken without moving the tripod. Finally, condition 9, the singleton object view, was generated by closely cropping the condition 6 image in Photoshop. Images in conditions 3 and 4 (large and small objects respectively) were programmatically generated by randomly 6 objects drawing from a database of large and small objects, and placing them in 3-across by 2-down grid. Images for condition 7 (vertical reachspaces) were selecting by finding reachable environments where the spatial layout of the objects was primarily on a vertical, rather than horizontal plane. This ranged from spaces with no horizontal extent (e.g. pegboard organization) to spaces with minimal horizontal extent (e.g. shelves). Finally, condition 8 and 10 images (regular reachspaces and scenes) were selected according to the same criteria as E1.

The main experimental protocol for Experiment 3 consisted of a blocked design with the 10 image conditions described above. Each run contained 4 blocks per condition, with blocks lasting 8s and consisting of 7 unique images and 1 repeated image. Within a block, each image was presented in isolation on a uniform gray background for 800ms followed by a 200ms blank. There were eight 10s fixation blocks

interleaved throughout the experiment, and each run started and ended with a 10s fixation block. A single run lasted 7 min (210 volumes). Participants viewed four runs of the experimental protocol. Participants' task was to detect a image repeated back-to-back, which happened once per block.

ROI definition. Experiment 3 used the same subjects and ROIs as Experiment 2.

Analysis. Responses across the 10 conditions were extracted from all object, scene, and reachspace ROIs. These responses were first visualized in a fingerprint profile. Next, to assess whether object, scene and reachspace ROIs had significantly different response profiles, we performed an analysis of variance to compare ROI types. Responses across the 10 conditions were averaged for all reachspace ROIs, scene ROIs and objects ROIs. These three response profiles were then submitted to a 2-way, condition-by-ROI type ANOVA.