



The contribution of object size, manipulability, and stability on neural responses to inanimate objects

Caterina Magri^{a,1,*}, Talia Konkle^a, Alfonso Caramazza^{a,b}

^a Department of Psychology, Harvard University, Cambridge, MA 02138, USA

^b Center for Mind/Brain Sciences (CIMEC), University of Trento, 38068 Rovereto (TN), Italy

ARTICLE INFO

Keywords:

Object representation
occipitotemporal cortex
functional magnetic resonance imaging,
real-world size

ABSTRACT

In human occipitotemporal cortex, brain responses to depicted inanimate objects have a large-scale organization by real-world object size. Critically, the size of objects in the world is systematically related to behaviorally-relevant properties: small objects are often grasped and manipulated (e.g., forks), while large objects tend to be less motor-relevant (e.g., tables), though this relationship does not always have to be true (e.g., picture frames and wheelbarrows). To determine how these two dimensions interact, we measured brain activity with functional magnetic resonance imaging while participants viewed a stimulus set of small and large objects with either low or high motor-relevance. The results revealed that the size organization was evident for objects with both low and high motor-relevance; further, a motor-relevance map was also evident across both large and small objects. Targeted contrasts revealed that typical combinations (small motor-relevant vs. large non-motor-relevant) yielded more robust topographies than the atypical covariance contrast (small non-motor-relevant vs. large motor-relevant). In subsequent exploratory analyses, a factor analysis revealed that the construct of motor-relevance was better explained by two underlying factors: one more related to manipulability, and the other to whether an object moves or is stable. The factor related to manipulability better explained responses in lateral small-object preferring regions, while the factor related to object stability (lack of movement) better explained responses in ventromedial large-object preferring regions. Taken together, these results reveal that the structure of neural responses to objects of different sizes further reflect behavior-relevant properties of manipulability and stability, and contribute to a deeper understanding of some of the factors that help the large-scale organization of object representation in high-level visual cortex.

1. Introduction

To recognize objects in the real world, large portions of the brain are utilized, including the ventral visual stream (Goodale and Milner, 1992; Ishai et al., 1999). Within this cortex there is a highly consistent organization of neural response tuning, with a tripartite large-scale topography distinguishing responses to animate entities, large inanimate objects, and small inanimate objects (Chao et al., 1999; Konkle and Caramazza, 2013; Proklova et al., 2016; Julian et al., 2017; Grill-Spector and Weiner, 2014). Some proposals have raised the possibility that the large-scale divisions in this cortex arise to support different behavioral needs (e.g., manipulating objects vs navigating in environments), linked to distinct underlying long-range brain networks (e.g., Konkle and Caramazza, 2017; Mahon and Caramazza, 2011). Thus, in the present work we focused on the inanimate domain, examining how both the size of

an object and its motor-relevance can explain brain responses to object categories.

Intuitively, the relationship between an object's size and the degree of interaction it affords in the world is evident: small objects tend to be graspable and easy to hold up, while larger objects tend to be used as support or for navigation. However, other relationships are possible: some small objects (e.g., picture frames) are not usually manipulated, while some large objects (e.g., pianos) are highly interactive. Here, we operationalize the dimension dealing with object interaction as *motor-relevance*. While similar conceptually to the property of manipulability that has been investigated in the past (Boronat et al., 2005; Campanella et al., 2010; Kalénine and Buxbaum, 2016; Kellenbach et al., 2003; Mahon et al., 2007), motor-relevance spans a wider definition in that it includes actions performed not only with the hands, but with other parts of the body as well. With this broader conceptualization, size can more easily dissociate from degree of interaction: large objects like swing sets and wheelbarrows tend to involve movements beyond hand-performed

* Corresponding author.

E-mail address: cmagri1@jhu.edu (C. Magri).

¹ Present address: Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218.

actions, while for small objects such as picture frames and smoke alarms there is no clear motor interaction associated.

To date, there is substantial evidence that motor-relevance is an important construct for visual object responses, particularly along the lateral aspect of the occipitotemporal cortex, a region known to show a preference for small inanimate over large inanimate objects (Konkle and Oliva, 2012; Konkle and Caramazza, 2013). For example, in this territory, there is a region with preference for tools (Chao et al., 1999; Bracci et al., 2012; Gallivan et al., 2013; Chen et al., 2016) which is closely overlapping with effector-specific regions (Bracci et al., 2012), and persists in the congenitally blind (Peelen et al., 2013; Wang et al., 2015). Further, this cortex shows connectivity with frontoparietal networks supporting actions (Simmons and Martin, 2012; Bi et al., 2015; Konkle and Caramazza, 2017), and is causally involved in tool-action discrimination judgments (Perini et al., 2014). While past studies have typically focused on objects that require hand-performed actions, the present study allows us to test whether the inclusion of non-hand-performed actions through the motor-relevance dimension would further explain neural responses and extend them to large objects.

In the ventromedial temporal cortex, another tool-preferring region has been observed by some studies in the medial fusiform gyrus when contrasting tools to animals (Chao et al., 1999; Garcea and Mahon, 2014; Mahon et al., 2007; Whatmough et al., 2002). However, the property of motor-relevance in this region is less clear: for example, responses to manipulable tools are not higher than to other inanimate objects (Chen et al., 2018; Mahon et al., 2007), though there is still some evidence for sensitivity to tools in neural adaptation signals (Mahon et al., 2007). Further, research focused on scene understanding has characterized these ventral-medial object responses along other kinds of object properties that are less related to motor-relevance. For example, responses in ventromedial scene-related regions are best predicted by objects invoking a local space that tend to stay fixed in the world, and are useful as landmarks (Troiani et al., 2014; Mullally and Maguire, 2011; Julian et al., 2017; Auger et al., 2013); see also (Epstein et al., 2014). These same areas are also known to show a preference for large over small inanimate objects (Konkle and Oliva, 2012; Konkle and Caramazza, 2013) and, interestingly, also to the names of such objects in both blind and sighted individuals (He et al., 2013).

Given these past observations, in the present study we aimed to clarify the degree to which brain responses along the occipitotemporal cortex are sensitive to motor-relevance distinctions, object size distinctions, and interactions between these interrelated properties. To do so, we designed our experiment to probe this question at two different levels of granularity. At a coarse level, we compared the broad relationship among object size and motor-relevance properties along a 2×2 design (large vs small, motor-relevant vs non-motor relevant objects), finding effects for both variables independently. Second, we included a nested factor in our design that enabled us to examine the same brain responses, but at the level of object categories (72 categories total). In this finer-grained level, an exploratory factor analysis revealed that three dimensions that respectively relate to object size, manipulability, and stability interact to account for the structure of brain responses in different regions. Finally, we consider the relative merits of feature- vs domain-based properties (e.g., manipulability, navigation relevance) as driving principles in the organization of high-level visual cortex.

2. METHODS

Data and code supporting the findings of this study are openly available as an Open Science Framework repository at <https://doi.org/10.17605/OSF.IO/U8XFK>.

2.1. Participants

For the fMRI experiment, 21 healthy adults with normal or corrected-to-normal vision participated in a 2-hour fMRI session. Three were later

removed for excessive head movement, and the remaining 18 subjects were analyzed (age: 18–30 years; 11 females). Each subject gave informed consent according to procedures approved by the Institutional Review Board at the University of Trento.

Across all of the online behavioral rating experiments there were 656 total participants. To develop and validate the stimulus set, 476 participants rated the dimensions of motor-relevance ($n=168$), size ($n=168$) and recognition ($n=140$) on Mechanical Turk. The varying numbers of participants were related to technical issues with launching HITs, and we additionally had to drop 7 raters from the motor-relevance and 5 raters from the recognition ratings. The remaining subjects ($n=168$ for size, 161 for motor-relevance, 135 for recognition) were analyzed. We did not collect more than one dimension from the same participant to avoid interference between dimensions. Additionally, to obtain ratings over the dimensions of manipulability, spatial definition, position variability, motion-based identity and interaction envelope, we obtained data from 36 participants for each dimension. Again, these participants rated only one dimension.

2.2. Stimuli

The stimulus set consisted of images from 72 object categories, where each object category belonged to one of four conditions (small motor-relevant, small non-motor-relevant, large motor-relevant and large non-motor-relevant) with 18 object categories per condition. Each object category included images of six exemplars of that category, for a total of 432 pictures in the stimulus set (Fig. 1a). Each image depicted an isolated object on a white background, and the image was sized to 512×512 pixels with the object's longest axis reaching the border of the image.

2.2.1. Stimuli validation

To develop and validate the stimulus set, we performed three separate behavioral studies using Amazon Mechanical Turk on an initial larger set of 20 object categories per condition with 7 unique exemplars for each object category (560 total images). Our aim was to construct a stimulus set in which items were recognizable, and where the size and motor-relevance dimensions were balanced.

- a *Recognition ratings.* Observers were presented with an item and reported whether they were familiar with this object (yes/no), and how easily they recognized the object on a scale from 1 (very hard to recognize) to 5 (very easy to recognize).
- b *Motor-relevance ratings.* Participants were presented with an item and rated the degree to which the object made them think of moving their hands or other parts of their body, using a scale from 1 (not at all) to 7 (very strongly).
- c *Size Ratings.* Participants were presented with an item and judged the size of an object as they would encounter it in everyday life, on a scale from 1 to 7. To provide a reference for the scale, pictures of a wristwatch (small size), a hiking backpack (medium size) and a bed (large size) were provided beside the values of 1, 4 and 7 on the scale.

We divided this initial dataset into 28 separate Mechanical Turk hits comprising 20 trials items (5 items from each of the four conditions). Only one exemplar for an object category was presented to each participant. Ratings were obtained for all 560 objects combining data from all participants, so that each image was rated by 6 observers along the measures of size and motor-relevance and by five observers along the measure of recognition.

In order to eliminate those items that were least recognizable, we dropped one item from each object category based on the familiarity score in the recognition experiment. We next eliminated two object categories that scored low in familiarity in order to balance the ratings from the other two experiments (Motor-relevance and Size experiments). This resulted in a new stimulus set with 18 objects per condition (72 total)

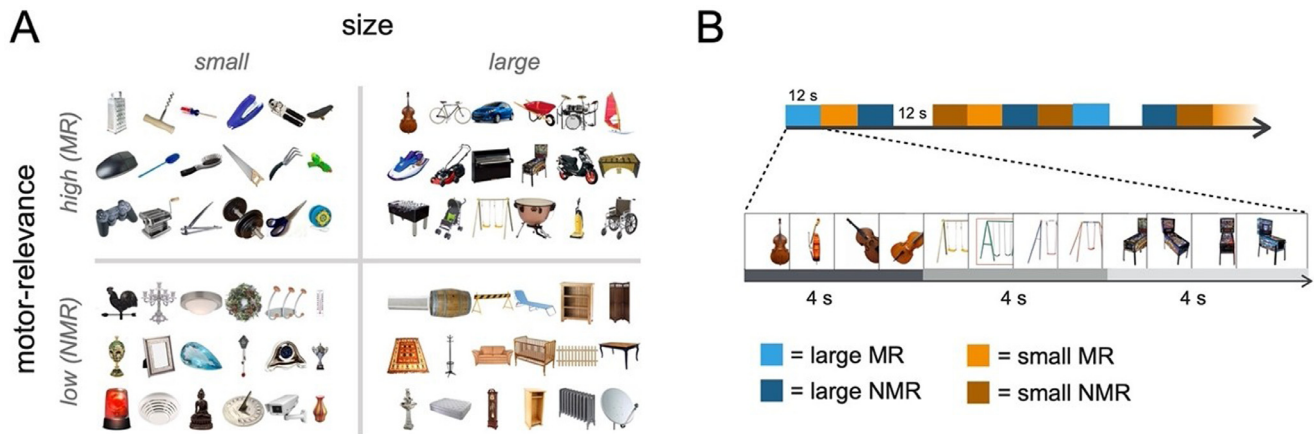


Fig. 1. a) Stimulus set: 72 inanimate object categories divided into four conditions (small motor-relevant, small non-motor-relevant, large motor-relevant, large non-motor-relevant). Each object category included six individual items, for a total of 432 individual images. b) fMRI task: subjects were presented with four consecutive pictures of the same object category in a 4s object category mini-block. Three 4s category-level mini-blocks (e.g., cello mini-block, swing set mini-block, pinball machine mini-block) were presented sequentially in a 12s condition-level main block (e.g., large motor-relevant objects block). The task for participants was to detect when a red frame surrounded one of the objects, which happened once per 12s block.

and 6 items per object. All results and analysis that follow refer to this new stimulus set.

To ensure these factors were balanced, we conducted several statistical tests. As expected, motor-relevant objects had higher motor-relevance scores than non-motor-relevant objects ($t_{35}=32.93$, $p<0.001$), and large objects had higher size ratings than small objects ($t_{35}=19.82$, $p<0.001$). Further, large and small objects did not differ in their average motor-relevance scores ($t_{35}=1.63$, $p=0.11$). The same was true when looking separately at large motor-relevant and small motor-relevant objects ($t_{17}=1.05$, $p=0.31$), and small non-motor-relevant and large non-motor-relevant object categories ($t_{35}=1.35$, $p=0.2$). Similarly, motor-relevant and non-motor-relevant objects did not differ in their average size score ($t_{35}=-0.72$, $p=0.5$) and neither did large motor-relevant and large non-motor-relevant relevant objects ($t_{17}=0.96$, $p=0.35$), nor small motor-relevant and small non-motor-relevant objects ($t_{17}=-1.76$, $p=0.1$).

We also ran a two-way ANOVA with size and motor-relevance as factors which replicated the t-test results, with no effect of motor-relevance ($F_1=0.44$, $p=0.509$) and significant effect of size ($F_1=0.497$, $p<0.001$) for the size ratings; no effect of size ($F_1=2.37$, $p=0.128$) and significant effect of motor-relevance ($F_1=846.92$, $p<0.001$) for the motor-relevance ratings, and no effect of either size ($F_1=0.87$, $p=0.353$) nor motor-relevance ($F_1=2.78$, $p=0.099$) for the recognition ratings.

Finally, we assessed the reliability of the ratings over these properties of size and motor-relevance in the final dataset. To do so, repeatedly dividing subjects in two groups, averaging ratings over object categories, and correlating them between these two halves of the data (5000 splits). We then averaged these split-half correlations over iterations. This analysis revealed that, even if there are some individual differences, at a group level with this sample size ($N>100$ per dimension), the ratings are extremely reliable: motor-relevance: $r = 0.97$; size: $r = 0.99$.

2.3. Image acquisition

Imaging data were acquired using a BioSpin MedSpec 4T scanner (Bruker) with an eight-channel coil. Functional data were obtained with an echo-planar 2D imaging sequence (repetition time TR: 2000ms; echo time TE: 33ms; flip angle: 73°; slice thickness: 3 mm; gap: 0.99 mm, with 3×3 in-plane resolution and 34 slices). Volumes were acquired in ascending interleaved order of slice acquisition.

2.4. Experimental design and statistical analysis

2.4.1. Design and procedure

During the fMRI experiment subjects viewed images of the small motor-relevant, small non-motor-relevant, large motor-relevant, large non-motor-relevant objects in a classic blocked design. In each run, stimuli from each of the 4 conditions were repeated 6 times each, in 12s stimulus blocks, with 6 interspersed rest blocks of 12s. In each run, the conditions were pseudo-randomly distributed, such that all 4 conditions plus a rest block appeared in a random order, and then were presented again in a new random order, and so on for all 6 repetitions. Images were each presented for 800ms with a 200ms blank between each stimulus. Fixation blocks also appeared at the beginning and end of each run, for 4 and 12 seconds respectively. Participants were instructed to maintain fixation and to press a button when a red frame surrounded one of the objects, which happened once per block. The full experiment consisted of six runs each lasting 6 minutes and 12 seconds.

Additionally, we added sub-structure in each stimulus block, to support more exploratory analyses about the relationship between all 72 object categories (see Fig. 1b). Specifically, each stimulus block was comprised of 4s mini-blocks. In each mini-block, three different exemplars from the same category were shown. For example, in a 12s block of large motor-relevant objects, a participant would see 4 different cellos, followed by 4 different swings, followed by 4 different pinball machines. Within each run all 72 object categories were presented exactly one time. The specific 4 exemplars that were displayed in each mini-block were randomly chosen without repetition from the set of 6 possible exemplars from that object category.

Each image was presented in isolation at a $\sim 8 \times \sim 8^\circ$ visual angle. Stimulus presentation was performed using the Psychophysics Toolbox package (Brainard, 1997) in MATLAB.

2.4.2. Category-preferring regions-of-interest

Category-preferring regions-of-interest (ROIs) were selected based on independent Functional Localizer runs, with the goal of identifying regions which have been widely investigated in past literature and known to process preferentially stimuli from specific categories over others.

These Functional Localizer runs consisted of a blocked design, in which each block included one of the following stimuli: 1) objects 2) scrambled objects 3) scenes 4) bodies 5) hands. Each run lasted 6 minutes and 36 seconds, with 50 6-second stimulus blocks (10 per condition) and ten 8-second rest blocks interspersed. In each stimulus-block,

six different stimuli from the same condition were presented, where each image was shown for 800ms followed by 200ms of fixation. Additionally, there was an orthogonal motion manipulation: for each of the 5 stimulus categories, in half of the blocks the object was presented at the center of the screen, and in the other half the object moved from the center out in one of 8 randomly directions randomly. This orthogonal motion manipulation was built into our default localizer scan protocol, and was independent of the research question at hand; thus, this manipulation was ignored in the present study. Observers were instructed to maintain fixation throughout the experiment and press the button when an exact image repeated back-to-back.

Two of the 18 subjects performed 8 experimental runs (instead of 6) and only 2 Localizer runs (instead of 3). Data from all experimental runs for these subjects were used for data analysis. Due to a technical error, localizer runs for one subject did not include scrambled objects. This subject was still included in all analyses except for those focusing specifically on LOC.

For the purpose of the current project, we were only interested in identifying category-selective regions Lateral Occipital Complex (LOC; Objects > Scramble; Grill-Spector, 2003), parahippocampal place area (PPA; Scenes > Objects; Epstein and Kanwisher, 1998) and Occipital Place Area (OPA; Scenes > Objects; Dilks et al., 2013). We were not able to localize area LOC in the subject for whom the scrambled objects condition was missing. To compare the coordinates of our contrasts of interest with the location of well-known category-selective regions, we also localized extrastriate body area (EBA; Downing et al., 2001) and a nearby hand-selective area (Bracci et al., 2012). All regions were identified at the individual level.

2.4.3. fMRI data analysis

Functional neuroimaging data were analyzed using SPM12 (Ashburner et al., 2014), MARSBAR (Brett et al., 2002) and bspmview (<https://www.bobspond.com/software/bspmview/>) on MATLAB (Versions 2014b and 2016a,b). Barplots and scatterplots were displayed using ggplot2 (Wickham, 2009) in RStudio.

The raw functional images were submitted to preprocessing, where the first 4 volumes were discarded from each run, slice scan-time correction was performed, followed by 3D motion correction, normalization and spatial smoothing (8 mm FWHM kernel). Data were modeled using standard general linear models (GLM). The first GLM included regressors for each of the four main conditions, with run regressors and motion correction parameters included as nuisance factors. The second GLM modeled each object category as a separate condition for a total of 72 regressors, with run regressors and motion correction parameters included as nuisance factors.

2.4.4. Size-preferring regions-of-interest

While we individuated category-specific ROIs through independent Functional Localizer runs, we had additional interest in examining regions which showed preferred response for small versus big inanimate objects. To this aim, we individuated small-object preferring and large-object preferring regions of interest using a subset of experimental runs independent from the ones that were analyzed. All regions were identified at the individual level. Spherical ROIs were defined around the individual peaks of activation for the whole-brain size contrast Small > Large (collapsing over motor-relevance) from two experimental runs. Univariate responses were extracted from the remaining experimental runs for size and motor-relevance conditions. These ROIs were defined with a 12mm sphere (257 voxels) centered around the peak positive and negative voxels in individual subjects, and then intersected with individual whole-brain gray matter masks. The selected peaks are referred to following the conventions of Konkle and Oliva (2010) and included small-object preferring bilateral inferior temporal gyrus (Small-ITG, left hemisphere [LH] $n=15$, right hemisphere [RH] $n=17$), and large-object preferring bilateral PHC (Large-PHC, LH $n=16$, RH $n=15$). After intersecting the 12mm ROI sphere with each subject's gray-matter mask, the

ROIs slightly varied in size across subjects. After intersection, the average size was 244 voxels for Small-ITG and 257 voxels for Large-PHC.

We further identified large-object preferring TOS (Large-TOS) based on the size contrast, and a series of category-selective regions using the Functional Localizer runs (see 2.4.2 for details). The regions identified with the Localizer were object-selective lateral occipitotemporal cortex (LOC), scene-selective parahippocampal place area (PPA), and scene-selective occipital-place area (OPA). Analysis and results for Large-TOS and the category-selective ROIs are reported in the Supplementary Material.

2.4.5. Voxel mask

We produced a mask of the most reliable object-selective voxels using a reliability-based voxel selection method developed for condition-rich designs (Tarhan and Konkle, 2020). To overview, this method identifies voxels whose response profile over conditions (72 in our case) is reliable across odd and even runs, and allows researcher to select a threshold of how reliable a voxel should be to be included, based on the multi-voxel pattern reliability of the conditions. The steps are as follows.

First, for each voxel, the group average beta values of 72 conditions were estimated from both odd runs and even runs, and correlated together, yielding a voxel-wise reliability map. Then, a given threshold (e.g., a voxel-wise split half reliability of $r>0.25$) yields a subset of voxels with reliability exceeding that threshold. Within this set of voxels, we then compute an estimate of how reliable the *multi-voxel patterns* are for each condition. That is, for each condition, the set of betas in these selected voxels were extracted for odd and even runs, and correlated together, to estimate the condition's pattern reliability; these estimates were averaged across the 72 conditions, yielding a final estimate of how stable the pattern-level responses are in that selected subset of voxels. This procedure was repeated for a range of voxel-wise reliability thresholds (range: $r>-1$ to $r>.95$, in steps of 0.05). The resulting condition-pattern reliabilities were plotted as a function of the voxel-wise reliability threshold: as a general rule, the higher the reliability threshold, the higher the condition multi-voxel reliability up to the point of plateau. The threshold level at which condition-to-condition correlations plateaued ($r=0.25$ in our case) was selected as the threshold for producing a reliability voxel mask.

2.4.6. Whole-brain conjunction analysis

To explore the size organization across motor-relevance, we performed a random-effects conjunction analysis at the individual level between small motor-relevant > large motor-relevant and small non-motor-relevant > large non-motor-relevant.

For each individual, at each voxel, t -values for the two contrasts were compared.

Given that we were interested in identifying voxels in which both contrasts were valid, the β -weight corresponding to the contrast with t -value closest to zero was assigned to the given voxel, obtaining a beta map of the weakest between the two contrasts for any given voxel.

This resulted in eighteen conjunction maps (one for each participant) which were then submitted to a one-sample t -test, and the resulting t -values were assigned to the corresponding voxels on a new map. The voxels with largest absolute values in these new t -maps are those that are strongly activated or deactivated by both contrasts in most subjects. The resulting maps were thresholded with $q(\text{FDR})<0.05$. Similarly, to explore the motor-relevance organization across size, we performed the same analysis but considering small motor-relevant > small non-motor-relevant and large motor-relevant > large non-motor-relevant contrasts.

2.4.7. Typical-Atypical contrast analysis

To further explore the relationship between size and motor-relevance, we compared maps between two contrasts: a *typical contrast* [small motor-relevant > large non-motor-relevant], which follows the real-world covariation of small objects being more motor-relevant than large objects; and an *atypical contrast* [small non-motor-relevant >

large motor-relevant]. Individual maps for each contrast were submitted to a second-level analysis to obtain a group map, which was FDR-thresholded at $q < 0.05$.

To assess the reliability of the two contrasts, we further produced individual maps for each contrast separately for odd and even runs and correlated them. We then submitted individual correlations for the two separate contrasts to a t-test to assess whether there was a significant difference in reliability between the maps produced by the two contrasts.

2.5. Exploratory analysis

2.5.1. Object property ratings

Ratings for the 72 objects in our dataset were collected on a series of other dimensions to identify the ones that were best associated with the response profiles of our ROIs. All studies were conducted on Amazon Mechanical Turk. Each of the participants ($n=36$) rated one exemplar of each of the 72 object categories, with six participants rating the same exemplars within each category. The following object properties were rated:

- 1 **Manipulability.** Manipulability tracks the degree to which hand-interactions with objects are prominent. Participants were presented with an item and rated the degree to which the object made them think of moving their hands, using a scale from 1 (not at all) to 7 (very strongly). The question asked to participants was very similar to the one posed to collect the motor-relevance dimension, but here we focus on actions performed with the hands rather than with other parts of the body.
- 2 **Motion-based Identity.** The degree to which an object's parts are expected to move and change position might be a relevant property for representing objects. Participants were asked to rate on a scale from 1 to 7 the degree to which motion is important in determining the identity of the object.
- 3 **Spatial Definition.** This property is defined as “the degree to which objects evoke a sense of surrounding space” (Mullally and Maguire, 2011). Participants were asked to rate on a scale from 1 to 7 the degree to which a background is evoked when looking at a specific object.
- 4 **Position Variability.** Objects can be more or less likely to change location in the environment (i.e., a car is more likely to change location than a swing). Participants were asked to rate on a scale from 1 to 7 the degree to which an object is likely to change position in everyday life.
- 5 **Interaction Envelope.** “Interaction envelope” was operationalized as a measure of the amount of space needed to interact with the object (e.g., a hamburger needs to be manipulated with two hands that cover its whole surface, thus it has a larger interaction envelope than a coffee mug which just needs one hand on its handle to be manipulated), following Bainbridge and Oliva (2015). Participants had to judge on a scale from 0 to 2 how many hands are often used when interacting with an object.

2.5.2. Pairwise correlations

We compared the similarity across our seven dimensions (size, motor-relevance and the additional ones collected post-hoc), and the reliability of each of the rated dimensions, in the following way. First, we divided participants into two groups and computed the mean rating for each group and each object category separately for each dimension, resulting in two group vectors of 72 mean ratings for each dimension. To compute the reliability of a dimension, we correlated the two group vectors with each other. To compute the correlation between two different dimensions, we correlated each dimension's group vectors with the other dimension's group vectors, and averaged across the four combinations.

2.5.3. Factor analysis

To understand how our critical dimensions of size and motor-relevance related to the variety of stimulus properties from the subsequent stimulus norming studies, we conducted a maximum-likelihood factor analysis on all seven dimensions (size, motor-relevance, manipulability, motion-based identity, spatial definition, position variability and interaction envelope) with a “varimax” rotation of the coefficients. This allowed us to identify underlying dimensions that were independent from each other, thus addressing the problem of covariance among all our ratings. This factor analysis was implemented with the *factanal* function in R. To determine the number of meaningful factors to extract from our dimensions, we performed a “parallel” analysis (Horn, 1965). This procedure involves generating random data and submitting them to the same factor analysis, iterated 5000 times. The average of the eigenvalues resulting from the parallel random-data factor analyses is then compared to the factors from the observed data: if the average eigenvalues from the parallel factor is smaller than the one from the data factor, then the data factor is kept. This procedure was implemented with the *fa.parallel* function in R. The three resulting factors can be summarized as relating to real-world size and context (f_{size}), manipulability (f_{manip}) and degree of movement ($f_{\text{stability}}$). For the following ROI and searchlight analyses, for ease of interpretation, we flipped the direction of $f_{\text{stability}}$ so as to positively correlate with degree of stability.

2.5.4. ROI analysis with factors as predictors

Following the factor analysis on the behavioral ratings of the objects, we performed a linear model analysis to predict the size ROIs responses from the resulting factor scores and their interactions. To estimate the response profile of each ROI, β -weights for each of the 72 object categories were extracted from the GLMs of the main experiment runs. Note that this analysis was performed on runs separate from the ones used to produce the ROIs. These betas were averaged across voxels, resulting in a vector of 72 values reflecting the ROIs overall response activation to each category. The ROIs activation profiles were computed for each subject using subject-specific ROIs and GLMs.

We next sought to explain these ROI activation profiles with the factors from the factor analysis above by employing a linear modeling approach. All linear models to predict an ROI activation were implemented with the function *lm* in R. We first looked for hemispheric interactions by submitting our size ROIs to a model with the three factors, hemisphere and their interaction as coefficients. No hemispheric interaction significantly explained either Small-ITG or Large-PHC's activation; hemisphere was thus removed from the models.

Next, we examined which of three nested models best explained activation in our size ROIs by comparing their resulting adjusted R square ($\text{adj}R^2$). The three models were: one with the three factors as predictors, with no interaction terms [$f_{\text{size}} + f_{\text{manip}} + f_{\text{stability}}$]; a second model that also included the two-way interaction terms [$f_{\text{size}} * f_{\text{manip}} + f_{\text{size}} * f_{\text{stability}} + f_{\text{manip}} * f_{\text{stability}}$]; and a third model that also included the three-way interaction term [$f_{\text{size}} * f_{\text{manip}} * f_{\text{stability}}$]. The model with the highest $\text{adj}R^2$ was considered the one most fitting the region and the significant coefficients were explored. Given that $\text{adj}R^2$ gives a measure of goodness-of-fit while also correcting for the number of parameters in a model, this measure can be employed to compare models with and without interaction term. If two models had comparable $\text{adj}R^2$ (rounded at the second decimal), the simpler model was picked. The same analysis was also performed on Large-TOS and in category-selective regions defined from the Functional Localizer (LOC-sphere, PPA-sphere and OPA-sphere; see Supplementary Material).

When comparing the nested models for the highest $\text{adj} R^2$, we found that both Small-ITG and Large-PHC were best explained by a model with all three factors and their two-way interactions (Small-ITG: $\text{adj. } R^2 = 0.33$; $F_{6,65} = 6.90$, $p < 0.001$; Large-PHC: $\text{adj. } R^2 = 0.27$; $F_{6,65} = 5.36$, $p < 0.001$), as compared to the model with only main effects (Small-ITG: $\text{adj. } R^2 = 0.23$; Large-PHC: $\text{adj. } R^2 = 0.21$) and the

model with main effects, two- and three-way interactions (Small-ITG: adj. $R^2 = 0.32$; Large-PHC: adj. $R^2 = 0.26$).

To evaluate the quality of the fit of the different coefficients, we looked at the resulting associated t - and p -values. T -values were computed by dividing coefficients by their standard errors, while p -values tested the significance against the hypothesis of obtaining given t -values if the coefficients were not actually contributing in explaining the dependent variable.

2.5.5. Searchlight regression analysis

Following our findings in the spherical ROIs, we next explored which factors would best fit voxels throughout object-responsive cortex. To this end, we performed a searchlight analysis, exploring whether we could observe any foci of real-world and/or motor-relevance preferring tuning beyond the ROIs we tested. The searchlight analysis was restricted only to voxels in the reliability mask—our rationale for this choice is to prevent over interpreting rating-brain relationships where the cortical responses are not stable. Also note that, as this is an exploratory analysis, brain maps are left unthresholded.

The searchlight analysis was performed in the following way: within the reliable voxel mask, a 6mm sphere was drawn around a voxel, and all voxels falling within that sphere were considered part of the neighborhood for that voxel. This procedure was performed iteratively so that each voxel within the object-selective mask was at the center of the sphere. In each searchlight ROI, voxel activation for all 72 object categories was averaged across all voxels within the sphere, and the resulting vector used as the dependent variable for two models.

For each search sphere, we fit two models. The first model related size and manipulability, with f_{size} , f_{manip} , and their interaction [$f_{\text{size}} + f_{\text{manip}} + f_{\text{size}} * f_{\text{manip}}$], which we refer to as *manip-model*. the second model related size and stability, with f_{size} , $f_{\text{stability}}$ and their interaction [$f_{\text{size}} + f_{\text{stability}} + f_{\text{size}} * f_{\text{stability}}$], which we refer to as *stability-model*. For each given searchlight, adjusted R^2 values for these two models were extracted and compared, and the model with the highest positive value was selected and assigned to that searchlight's central voxel.

To visualize which models are fitting best and where, we looked separately at small-preferring and large-object preferring voxels. Small- and large-object preferring voxels were identified based on whether the coefficient for the main effect of f_{size} was positive (preference for small) or negative (preference for large). We then visualized the coefficient maps for the interaction terms $f_{\text{size}} * f_{\text{manip}}$ and $f_{\text{size}} * f_{\text{stability}}$ separately for large- and small-preferring voxels, comparing them with a map of the F -values for the winning model associated with each voxel.

Finally, to further explore the direction of the interaction, we visualized the activation of small-object preferring voxels and big-object preferring voxels with scatterplots. To select the voxels for the plots, ROIs were created by identifying the peak of activation (that is, the voxel with strongest positive or negative main effect of f_{size}), and intersecting a binary sphere (radius = 20mm) centering at that point with all the voxels that loaded on f_{size} towards the same direction.

3. Results

3.1. Whole-brain conjunction analysis

We first examined whether the large-scale neural organization by real-world size of inanimate objects is preserved when disrupting the natural covariance between size and motor-relevance. Note, here we refer to the “real-world size” of objects to refer to the physical size of the space the real object would occupy in the world (and not the physical size that the depicted object's image occupies as it is projected on the scanner screen). To do so, we conducted a conjunction analysis to isolate regions which show a large vs small object difference that holds when comparing motor-relevant as well as non-motor-relevant objects (see Methods).

When looking at the real-world size organization, we found regions that showed a preference for small over large objects, and for large over small objects, regardless of the motor-relevance preference. The whole brain maps can be visualized in Fig. 2a, with the group MNI coordinates at the bottom of the figure. With respect to small objects, we observed activation bilaterally in ITG and in the left hemisphere and along the lateral aspect of the fusiform gyrus. In the left hemisphere, the region for small objects expanded posteriorly to the lateral occipital cortex. With respect to large objects, along the ventral surface of the occipitotemporal cortex, we observed bilateral regions corresponding to the parahippocampal cortex. A preference for large objects was also present more dorsally in the lateral surface in the occipital place area (OPA) in individual participants, but this region did not survive the group-level conjunction test at the FDR threshold depicted in Fig. 2a. Overall, this group-level conjunction analysis reveals that, even when taking into account motor-relevance, the real-world size of objects drives large-scale differential responses across the ventral stream.

Next, we examined whether also motor-relevance, when disrupting its natural covariance with size, would elicit a large-scale organization across this cortex. Another whole-brain conjunction analysis was conducted, comparing motor-relevant objects vs non-motor-relevant objects, requiring this relationship to hold for both large object and small objects independently. The results are mapped in Fig. 2b, with the group MNI coordinates at the bottom of the figure. Along the lateral surface, we observed regions with a stronger response to motor-relevant objects, with peaks in LO bilaterally. Along the medial aspect of the ventral surface, we found stronger responses to non-motor-relevant objects primarily in the right hemisphere, with the peak activation located slightly anterior to the large-object preferring region. This analysis reveals that the dimension of motor-relevance also elicits stable response differences in parts of object-responsive cortex.

Notably, these conjunction maps yielded relatively similar organizations. Specifically, lateral regions prefer both small and motor-relevant objects; while ventromedial regions prefer both large and non-motor-relevant objects. These findings indicate that neither size alone nor motor-relevance alone are sufficient to explain the large-scale organization of responses to objects along occipitotemporal cortex. Further, these results highlight that there is a clear association between the two dimensions and how they map across the cortex.

3.2. Typical-atypical contrasts

To further explore the relationship between size and motor-relevance, we compared maps between two contrasts: a *typical contrast* (small motor-relevant > large non-motor-relevant), which follows the real-world covariation of small objects being more motor-relevant than large objects; and an *atypical contrast* (small non-motor-relevant > large motor-relevant; Fig. 3a). These maps are shown in Fig. 3.

When looking at surface maps produced by the typical contrast, we find extensive activations along the lateral and ventral surfaces (Fig. 3b). In contrast, the surface maps produced by the atypical contrast (Fig. 3c) are less extensive and also less reliable in a split-half analysis (typical split-half map correlations: $M=0.61$, $SD=0.18$; atypical: $M=0.43$, $SD=0.24$; $t_{17}=2.70$, $p<0.05$).

These topographic observations mirror the ecological covariation of these factors (Fig. 3a): while it is theoretically possible to artificially dissociate size and motor-relevance as we have done in this stimulus set, in every day experience small objects tend to be more motor-relevant and large objects less so. And we found this covariation to also be evident in the neural data (see Fig. 3b,c). The degree to which this is driven by the prescience of particular visual feature combinations, differences in object familiarity, or a combination of these factors, is left open in this analysis. However, overall, these data show the importance of the interaction between size and motor-relevance in driving more reliable and robust neural responses, thus revealing the importance of the natural covariation between these two properties in the visual system.

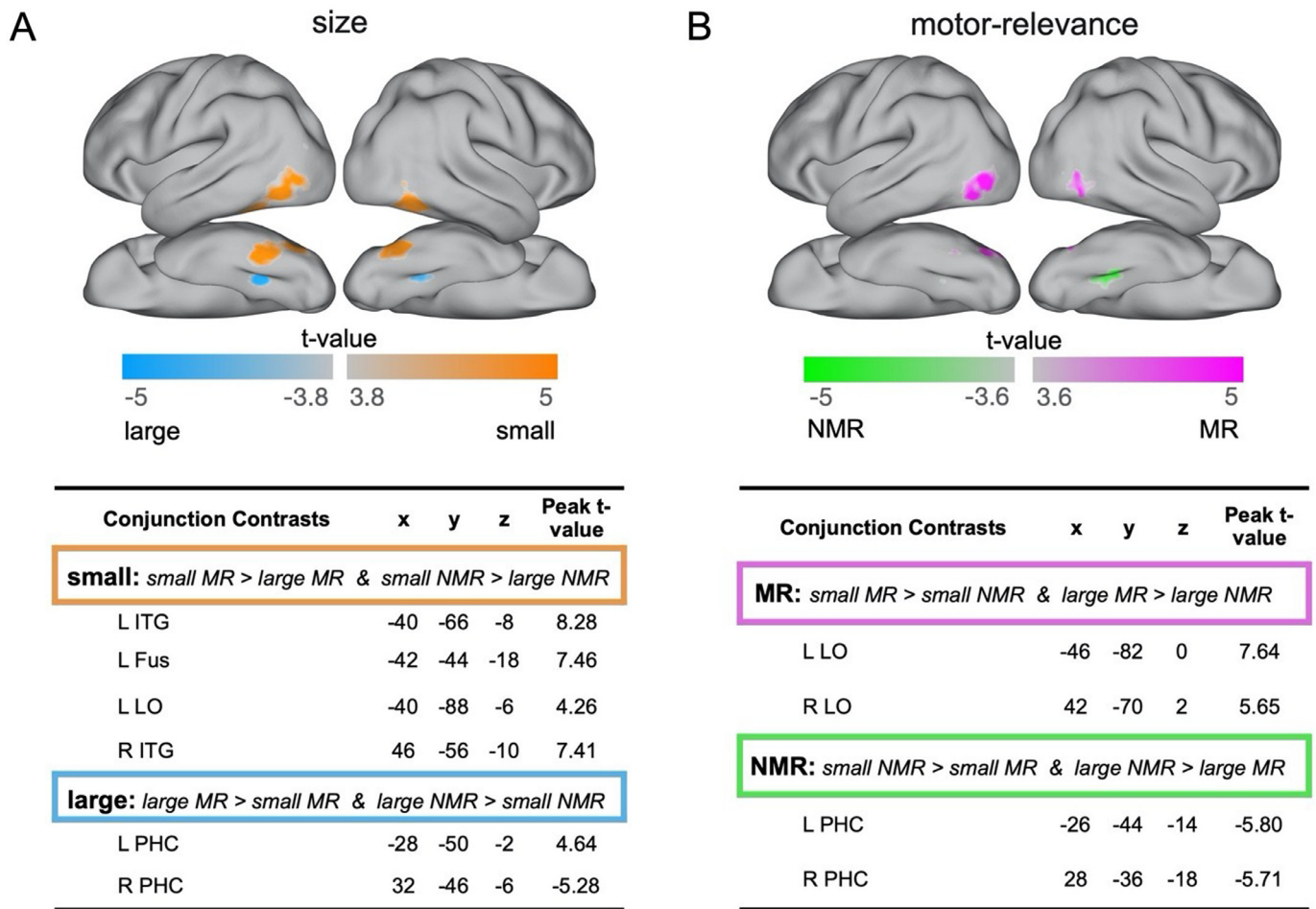


Fig. 2. Whole Brain Random-Effects Conjunction Analysis. a) Size conjunction contrasts, where both [small motor-relevant > large motor-relevant] and [small non-motor-relevant > large non-motor-relevant] relationships hold. a) Motor-relevance conjunction contrasts, where both [small motor-relevant > small non-motor-relevant] and [large motor-relevant > large non-motor-relevant] relationships hold. Contrasts were thresholded at $q(\text{FDR}) < 0.05$.

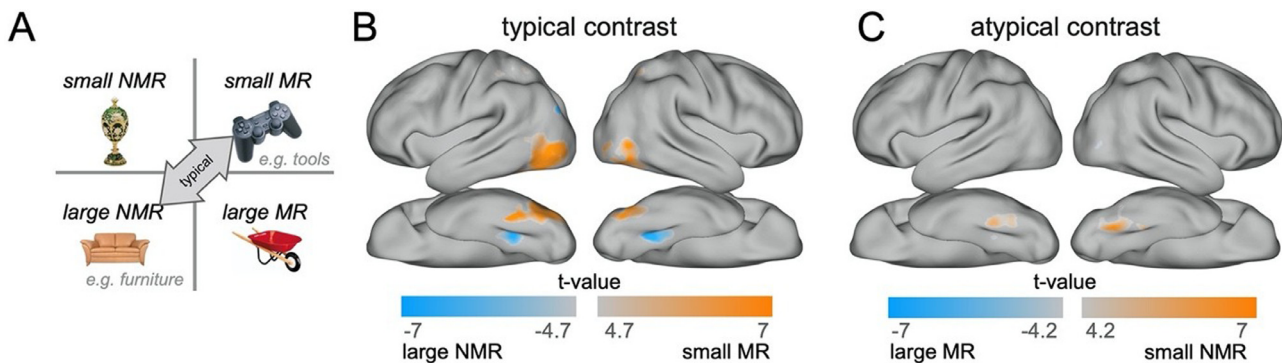


Fig. 3. Typical/Atypical contrast Analysis. a) simplified representation of the size \times motor-relevance design. Arrow points to the typical contrast conditions, representing the covariation commonly observed in the environment. b) Typical contrast: small motor-relevant > large non-motor-relevant. c) Atypical contrast: small non-motor-relevant > large motor-relevant. Contrasts were thresholded at $q(\text{FDR}) < 0.05$.

3.3. Exploratory analyses

3.3.1. Factor analysis

We next explored the possibility that the relationship between size and motor-relevance may be more multifaceted than what can be gathered from a simple 2×2 design. Our goal for this analysis was to understand and explore how other, related properties that have been proposed in the past relate to our findings and predict neural responses at

the category level. To do so, we collected ratings along five additional dimensions of our 72 object categories and submitted them to a factor analysis together with size and motor-relevance.

First, manipulability ratings were obtained, which focused on hand movements specifically, in order to understand the extent to which our motor-relevance dimension correlated with this previously studied dimension (Mahon et al., 2007). We also examined a number of dimensions that have been previously proposed as significant mod-

ulators of tool regions or scene regions, namely interaction envelope (Bainbridge and Oliva, 2015), spatial definition (Mullaly and Maguire, 2011), and motion-based identity (Beauchamp, 2005). Finally, based on our observation regarding low responses to vehicles in Large-PHC, we obtained “position variability” ratings, capturing how often an object changes its position in the environment. Note that these properties were defined after having seen the data, and thus any role they play in explaining neural data requires independent follow-up studies to confirm.

The relationship among these ratings and our original size and motor-relevant dimensions are detailed in Fig. 4. All ratings had a relatively high reliability (Fig. 4a, shaded diagonal numbers), with a range of similarity relationship between any pair of dimensions (Fig. 4b). There are a couple observations to note from inspecting these pairwise correlations and scatter plots. First, by design, size and motor-relevance were completely uncorrelated in this stimulus set. Second, the correlation between motor-relevance and manipulability was at ceiling, virtually identifying the same property. As mentioned in the Introduction, most motor-relevant objects involved hands for their interaction; including whole-body actions in the motor-relevance score had a minimal or non-existent effect (the only exception being the skateboard which was high in motor-relevance but low on manipulability).

Next, we extracted the latent dimensions within this dataset using factor analysis (see Methods). While these ratings were correlated with each other to varying degrees, this analysis allowed us to identify orthogonal, independent dimensions that summarize their underlying space. A parallel analysis procedure determined that three factors should be retained, which were sufficient to explain more than 80% of all variance among the ratings. The property loadings for each factor are shown to the right in Fig. 4a.

The first factor was loaded mostly by manipulability, motor-relevance and interaction envelope—it seemed thus to be most related to *actions and hand-object interactions*. The second factor was loaded by size and spatial definition, and was thus most related to the *physical presence* of the object. Finally, motion-based identity, position variability and motor-relevance loaded strongly on the third factor, which seemed thus to capture the *mobility/movability of the object*.

Considering these factors within the context of our main design, our motor-relevance dimension was effectively split into two factors, one related to hand-object interactions (f_{manip}) and one related to an object's mobility/movability ($f_{\text{stability}}$), and both of these were separate from the more physical properties of size and, to a lesser degree, spatial definition (f_{size}). Thus, while our stimulus set was collected as varying along only two primary factors (size and motor-relevance), the factor analysis indicated that this stimulus set is better characterized along three factors (size, manipulability and stability).

We next explored how well the three factors extracted from our ratings can explain the neural response profiles, both in a targeted ROI analysis and in a broader searchlight analysis. Note that these analyses were unplanned and should be considered post-hoc and exploratory in nature.

3.3.2. ROI analysis

To explore how well the three factors could explain fMRI responses, we employed a linear modeling analysis explaining the activation profile in each ROI, with the three factors obtained from the factor analysis and their interactions as regressors (see Methods). The goal was to see which weighted combination of factors best predicted a region's response variation to the 72 objects. This finer-grained analysis looking at responses to all 72 object categories was possible due to our nested fMRI protocol design (see Methods), in which the neural responses could be modeled at the category-level for each of the 72 categories independently (see Fig. 1). Note that these three factors were determined only from the behavioral ratings of our stimulus set. Further, for interpretive clarity, we reversed the sign of $f_{\text{stability}}$ so that high values indicate high “fixed-

ness” (rather than high mobility.) This transformation does not affect the significance of the statistical results.

After determining which model was a best fit for each region, we looked at the factors within the winning models that significantly contributed to the fit (see Methods), and whether they were positive or negative, to understand the models' relationship to the data (Fig. 5a).

In Small-ITG, the results indicate that the best prediction of neural response magnitude was given by a combination of size and of the *manipulability* component of motor-relevance: there was a significant effect of f_{size} , as well as the interaction of f_{size} and f_{manip} (see Fig. 5a, top for statistical results). The interaction between these factors further indicates that it is specifically small objects involved in hand-object actions which drive the strongest responses in this region. Thus, the distinction of hand-object action and manipulability is a better descriptor of the responses in this region over the more general concept of motor-relevance.

In Large-PHC we observed a different pattern: the coefficients that significantly explained the model were f_{size} and the interaction of f_{size} and $f_{\text{stability}}$ (see Fig. 5b, bottom row for statistical results). Thus, responses in this region were driven most by large objects that are stable in the environment. This result in Large-PHC clarifies what we observed in earlier analyses employing the 2×2 design: rather than the low-action component of motor-relevance, it might be the *fixedness* component of this construct the one most driving neural representations in this area. That is, the *stability* component of non-motor-relevance is the one to best predict response magnitude in Large-PHC.

To better understand these relationships, we visualize them in Fig. 5c: for each ROI, the activation to each category is plotted on the y-axis as a function of the manipulability factor for Small-ITG, and the stability factor for Large-PHC. Each dot is color-coded based on the f_{size} value for the corresponding object. This plot reveals that the manipulability factor drives activation in Small-ITG but mostly for small objects (increasing activity for more manipulable small objects, see Fig. 5c, top). In contrast, inspection of the scatterplot for Large-PHC suggests that activation is driven by stability but mostly for large objects (increasing activity for more stable large objects, see Fig. 5c, bottom). These relationships are also evident in the activation profiles of the two ROIs in Fig. 6: For Small-ITG, tools are the small objects that drive the most activation (e.g. nail clipper, hand rack, yo-yo; see Fig. 6a), while for Large-PHC, means of transport tend to be the large objects that drive the least activation (e.g. bike, scooter, jet-ski; see Fig. 6b).

Taken together, these analyses begin to characterize the nature of the ecological covariation between size and motor-relevance, clarifying how these object properties fit together. For example, no one factor alone (size, manipulability, stability) best accounted for the responses in any ROI. Rather, manipulability (or hand-relevance) of an object matters more for small objects in the lateral cortex, and stability of an object matters more for large objects in ventral cortex. Finally, these findings are consistent with prior literature that has focused on either the ventral or lateral surface separately (e.g., Bracci et al., 2012; Mullaly and Maguire, 2011), unifying these separate observations in one common data set and analysis procedure.

3.3.3. Additional ROIs

To more directly relate our findings to previous literature, we also conducted these same analyses in the classic scene regions, parahippocampal place area and occipital place area (PPA and OPA; Epstein et al., 2014), and object-responsive lateral occipitotemporal cortex (LOC; see Methods and Supplementary Material). These regions are adjacent and partially overlapping with the size ROIs (see also Konkle and Oliva, 2012). The findings in these classic regions generally converge with what was observed for the size ROIs: in LOC, there was a significant contribution of f_{size} , and a significant contribution of the interaction between f_{size} and f_{manip} . In PPA, there was a significant contribution of the size factor, and marginally of stability and of the interaction between size and stability (see Supplementary Material).

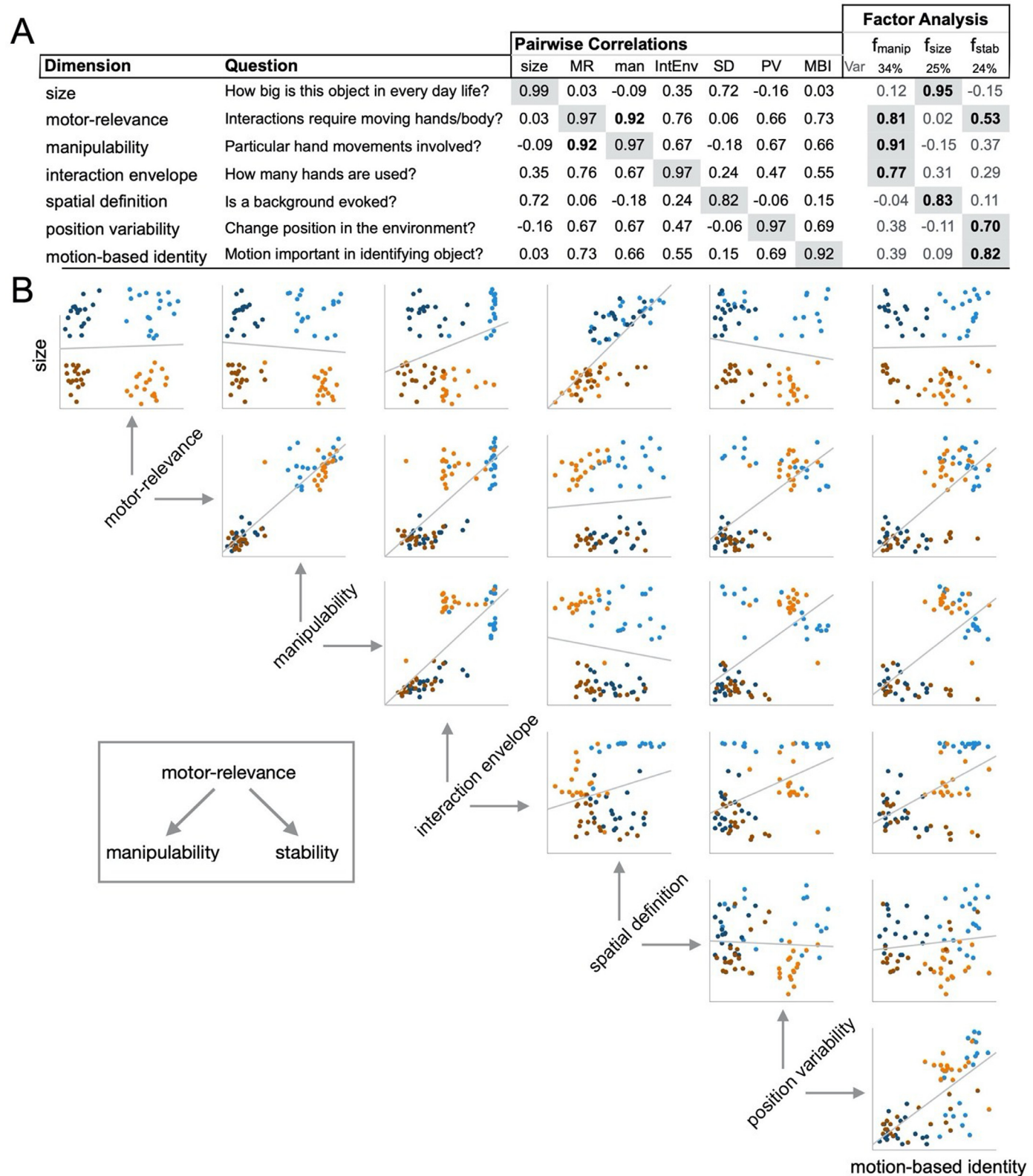


Fig. 4. Object dimension similarity and Factor Analysis. a) For each rated object property, the subplot shows the dimension label, the question asked to elicit ratings on the dimension, the correlation to the other dimensions, where the rating's reliability is shaded, and the dimension's loading in a factor analysis. b) Scatter plots are shown for all pairs of 7 dimensions, where each dot reflects an object category and is color-coded based on its size and motor-relevance rating (light orange – small motor-relevant; dark orange: small non-motor relevant; light blue: large motor-relevant; dark blue: large non-motor-relevant).

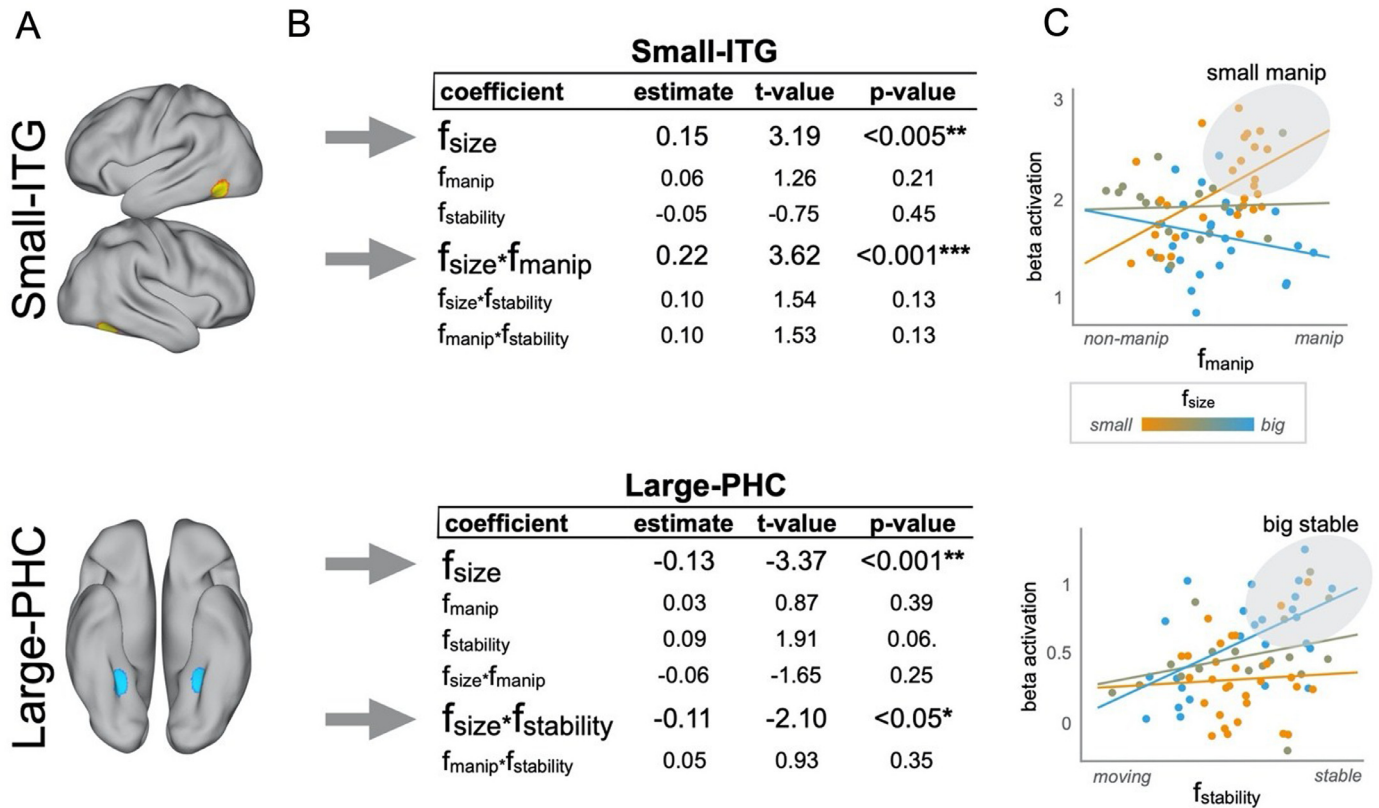


Fig. 5. ROI analysis. a) Visualization of size ROIs Small-ITG and Large-PHC in one example subject. b) GLMs with ROIs' activity as dependent variable and factor scores as independent variables. Shown are estimates for all predictors included in the model, t-value and p-value. The t-value was measured by dividing the coefficients by their standard errors. The p-value tested the hypothesis of obtaining the observed t-value if the coefficient were actually zero. c) Top: scatterplot of Small-ITG's activation by f_{manip} (x-axis) and f_{size} (y-axis). Dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size). Bottom: scatterplot of Large-PHC's activation by $f_{\text{stability}}$ (x-axis) and f_{size} (y-axis). Same color-coding scheme as described for the plot above.

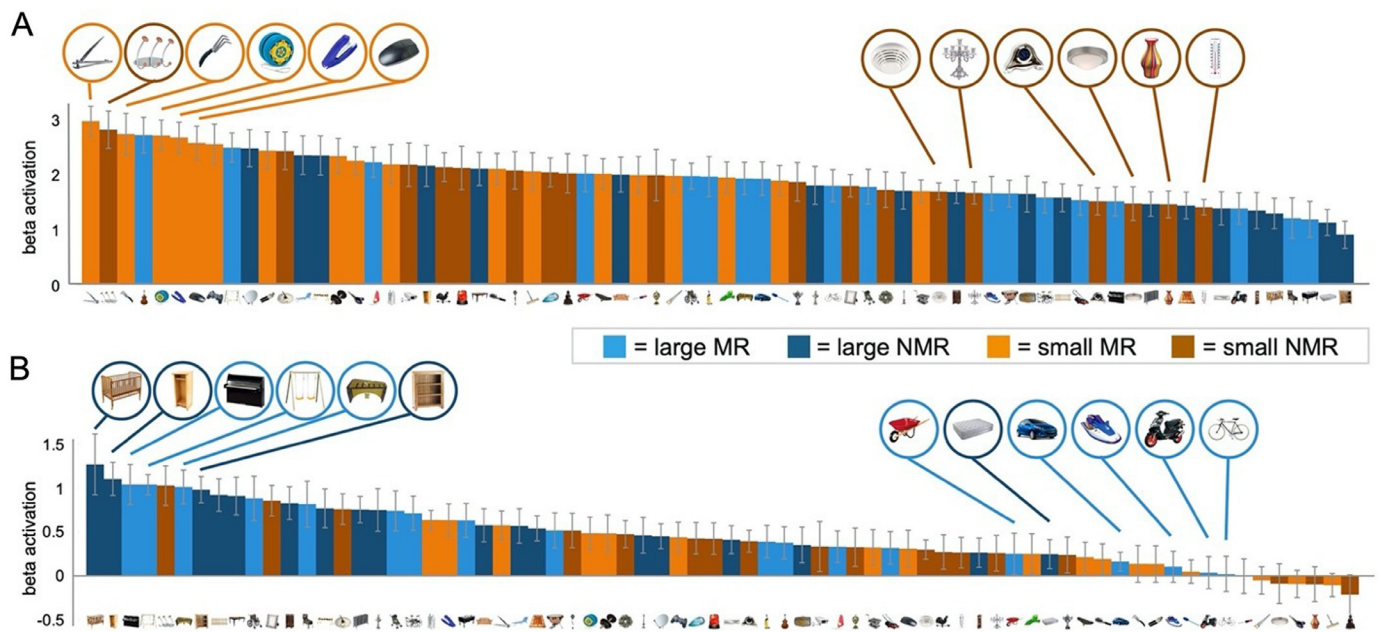


Fig. 6. Barplots of ROIs activation. Mean activation to each of the 72 object categories are plotted for each region (a: Small-ITG; b: Large-PHC), where the categories are rank-ordered from most to least active. Object categories are color-coded based on their size and motor-relevance conditions (light orange: small motor-relevant; dark orange: small non-motor-relevant; light blue: large motor-relevant; dark blue: large non-motor-relevant). Error bars reflect ± 1 SEM across subjects.

Further, we conducted these analyses in an additional occipitodorsal region identified with the [Large > Small] size contrast from two runs of the main experiment (Large-TOS) and in adjacent scene-selective occipital place area (OPA; see Supplementary Material). In Large-TOS, we found a result parallel to our observation in PPA: a significant contribution of f_{size} and of the interaction between f_{size} and $f_{\text{stability}}$. However, this result did not replicate in OPA where only f_{size} had a significant contribution.

3.3.4. Searchlight analysis

By focusing on the ROIs produced by the size contrast, we were able to explore whether other factors beyond size and motor-relevance were important in these ROIs—to which the answer is clearly yes: motor-relevance was indeed split into two sub-dimensions, one more related to manipulability, and one more related to stability, which dissociated in explaining small- and large-object ROIs. However, by performing a targeted ROI analysis, we might have missed additional regions with different associations between these factors; thus, we next performed searchlight analysis. Note, however, that this is a purely exploratory analysis, based on factors that were specified post-hoc (after neural data collection). Given the exploratory nature of this analysis, results are presented unthresholded, and interpreted with appropriate caution.

Given the complexity of the number of possible interactions among these factors, we approached the searchlight analysis with a more targeted comparison. Specifically, we mapped which regions showed a better model fit among two candidate models: one that associates size and manipulability (manip-model) and one that associates size and stability (stability-model; see Methods). All searchlight analyses were conducted within a reliable voxel mask (see Methods). The results of this analysis are shown in Fig. 7.

From this analysis, we can gather two main conclusions. First, we found that the voxels with best fitting models (i.e., the ones with highest F-value, see Fig. 7, center) focused mostly in the lateral occipitotemporal region and in the ventromedial occipitotemporal regions, in areas mostly overlapping with size ROIs Small-ITG and Large-PHC (outlined in black in Fig. 7a,b). Thus, our targeted ROIs seemed to cover most OTC areas where voxels' information could be explained by a combination of the three factors.

Second, we found that the best-fitting voxels resulting from the searchlight analysis largely confirmed our targeted ROI analyses. Indeed, small-object preferring voxels in the lateral occipitotemporal region (i.e., the ones with positive f_{size} coefficients for the winning model) presented a positive interaction of the size and manipulability factors (Fig. 7a top, colored in yellow), similarly to what was observed in the ROI analysis for Small-ITG: that is, the smaller f_{size} (the smaller the object), and the higher f_{manip} (the more manipulable the object), the stronger the activation. Similarly, we found that large-object preferring voxels in ventromedial occipitotemporal region (i.e., the ones with negative f_{size} coefficients for the winning model) presented a negative interaction of the size and stability factors (Fig. 7b, top, colored in green), similarly to what observed in our ROI analysis for Large-PHC: that is, the lower f_{size} (the bigger the object), and the higher $f_{\text{stability}}$ (the more stable the object), the stronger the activation. For ease of comparison, we have drawn a black outline on the searchlight coefficient maps of the size ROIs Small-ITG (Fig. 7a) and Large-PHC (Fig. 7b) for one example subject, from which it is clear that the targeted ROI analyses do in fact highlight the major structure in neural responses by our three factors.

At lower F-value levels, we observed two more nuanced relationships. However, given the lower fit of either manip-model or stability-model within these voxels, one can only draw tentative inferences from these results.

First, a significant portion of small-object preferring voxels in the lateral occipitotemporal surface (Fig. 7a, top, colored in cyan) are fit best by models with size and stability, and show a negative interaction between the size and stability factors. Inspection of activation to all 72

object categories in this area shows a preference for moving over non-moving things (see Supplementary material) which is in line with the presence of movement-preferring region MT+ in the lateral surface.

Second, we also observed large-object preferring voxels in the dorsal occipitoparietal region, and weakly in the superior/anterior parietal cortex, where most voxels (Fig. 7b, top, colored in yellow) indicate a positive interaction of the size and manipulability factors while a few voxels (colored in green) indicate a negative interaction of the size and stability factors. However, inspection of the scatter plots for these regions showed that these interactions were not as simple and easily interpretable as the ones observed in areas neighboring our size ROIs (see Supplementary Material). Further, note that these scatter plots are post-hoc explorations of a the pattern in these regions, and are not independent of the selection criteria. From these results, it is clear that there are systematically varying object responses in this dorsal occipital part of the cortex, but none of the three factors or their interactions cleanly provides a description of what explains the object activations.

4. Discussion

Here, we examined whether the large-scale organization of neural responses to object size could be better explained by the factor of motor-relevance. This was not the case – we found that the contrasts of small vs large objects elicited a clear size organization for both motor-relevant and non-motor relevant objects. We also found that motor-relevance, independent of size, accounted for some of the large-scale structure of neural responses. This observation confirms the importance of motor interaction as an explanatory dimension in the ventral visual stream. Crucially, however, further analyses showed the two dimensions to not be independent, in that more typical combinations (small and motor-relevant, large and non-motor-relevant) elicited more reliable representations than less typical ones.

Our subsequent exploratory analyses helped to refine and clarify these results. Specifically, a factor analysis revealed that the construct of motor-relevance could be divided into two underlying dimensions, one better related to *manipulability* (i.e., related to hand performed actions), and one better related to an object's *physical presence* and stability. Further ROI and searchlight exploratory analyses revealed that the manipulability factor was a better descriptor of activation in small-object preferring lateral OTC, while the stability factor was a better descriptor of activation in large-object preferring ventromedial OTC.

Broadly, these results begin to characterize how the ecological covariation between size and behaviorally relevant object properties contributes to the observed dissociations between object categories in the ventral stream regions with a preference for the inanimate domain. In the next two sections, we situate these findings in the literature, and discuss how they inform the deeper theoretical question of what drives the neural organization of inanimate objects.

4.1. Dissociations within the inanimate-domain

Objects are characterized by numerous visual properties, real-world size being one of them. However, size tends to correlate with other visual and non-visual properties. There are suites of partially related factors that predict neural responses to inanimate objects, with a major division between small, manipulable objects on one side and large, stable, navigationally relevant objects on the other (e.g., He et al., 2013; Peelen et al., 2013; Julian et al., 2017; Mullally and Maguire 2011; Troiani et al., 2014; Auger et al., 2013; Bracci et al., 2012; MacDonald and Culham, 2015; Bainbridge and Oliva, 2015). Overall, these findings point to a richer organization than simply object size as such, where different brain regions may be sensitive to size in the presence of other object properties for different reasons (Bi et al., 2015).

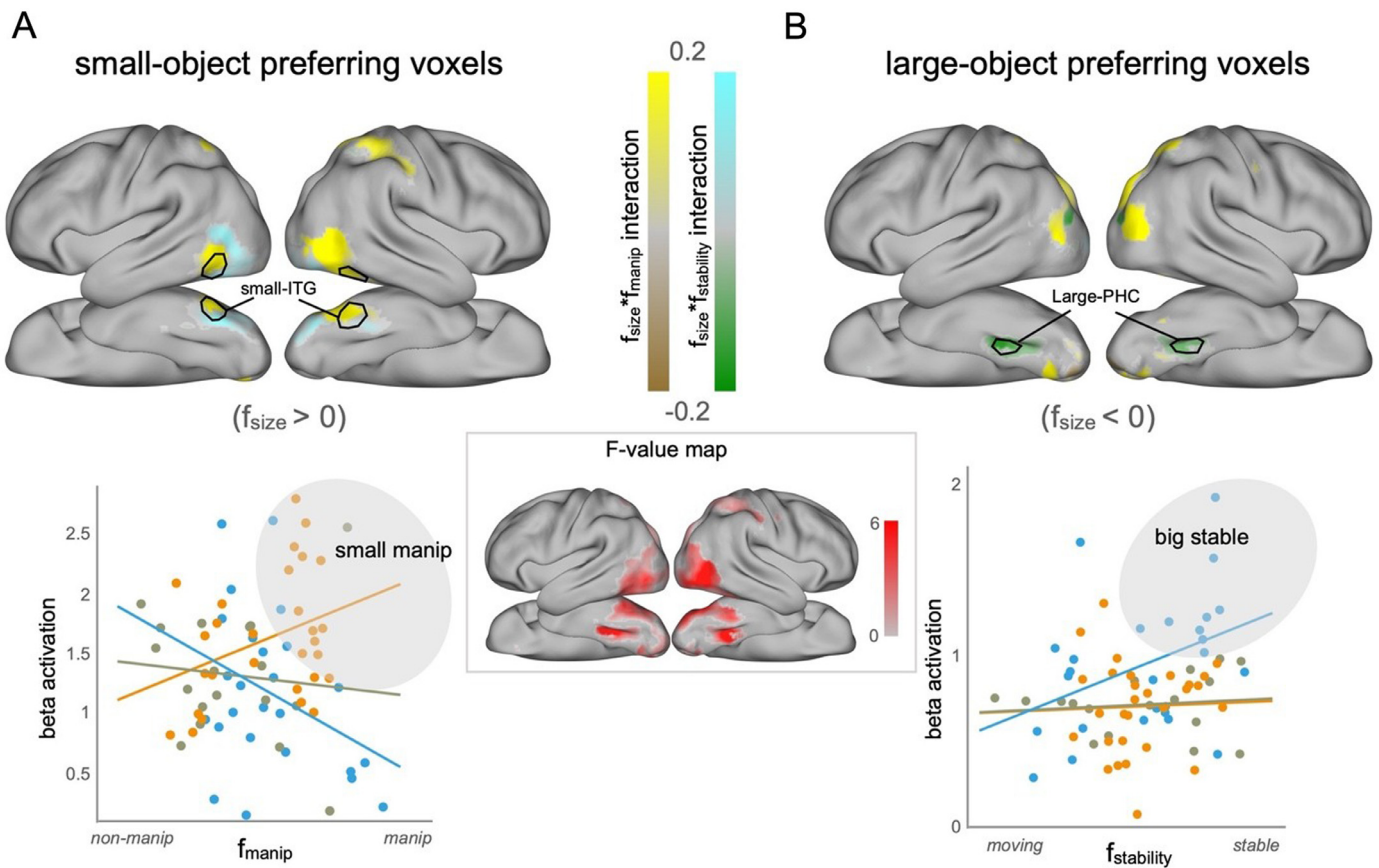


Fig. 7. Searchlight Analysis. a) Top: surface map of interaction coefficients for small-object preferring voxels (i.e., voxels with a positive f_{size} coefficient). In voxels in which the winning model was the manip-model, that voxel is colored according to its $f_{size} * f_{manip}$ interaction coefficient estimate in a brown to yellow scale. In voxels in which the winning model was the stability-model, that voxel is colored according to its $f_{size} * f_{stability}$ interaction coefficient estimate in a green-to-cyan scale. Bottom: scatterplot of activation by f_{manip} (x-axis) and f_{size} (y-axis) for all voxels showing a positive $f_{size} * f_{manip}$ interaction in lateral OTC. Each dot represents one of the 72 object categories from our stimulus set, and dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size). b) Top: surface map of interaction coefficients for large-object preferring voxels (i.e., voxels with a negative f_{size} coefficient). In voxels in which the winning model was the manip-model, that voxel is colored according to its $f_{size} * f_{manip}$ interaction coefficient term in a brown-to-yellow scale. In voxels in which the winning model was the stability-model, that voxel is colored according to its $f_{size} * f_{stability}$ interaction coefficient estimate in a green-to-cyan scale. Bottom: scatterplot of activation by $f_{stability}$ (x-axis) and f_{size} (y-axis) for all voxels showing a negative $f_{size} * f_{manip}$ interaction in ventromedial OTC. Each dot represents one of the 72 object categories from our stimulus set, and dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size). Note that the medial side of the brain is not depicted as no voxels on this facet of the brain were included in the set of reliable voxels, and that these scatterplots show data that is not independent of the selection criteria.

For example, along lateral OTC and middle temporal gyrus, there is a mosaic of partially overlapping regions whose response magnitudes are predicted by a range of related factors, from more primitive object shape (Grill-Spector, 2003), to real-world size (Julian et al., 2017; Konkle and Oliva, 2012), to tools and other objects that are manipulated by the hands (Lewis 2006; Bracci et al., 2012; MacDonald et al., 2015; Mruczek et al., 2013; Lingnau and Downing, 2015). These regions are also adjacent to motion-selective cortex and body-preferring regions (Grill-Spector and Weiner, 2014), including a region that jointly responds to both tools and hands, spanning the animate/inanimate divide (Bracci and Peelen, 2013; Striem-Amit et al., 2017). While the number of regions and exact functional divisions along this cortex are still being clarified, from our results and these others, it is clear that these regions are most strongly related to small, movable, manipulable items.

Along ventromedial OTC, where scene-preferring PPA is located, there are now a number of converging results characterizing the object properties that are best associated with responses to isolated objects. These factors include the degree to which the object invokes a local tridimensional space, its size, its tendency to stay fixed in the world, and can be generally summarized by an overarching construct of spatial stability and landmark suitability (Troiani et al., 2014;

Mullally and Maguire, 2011; Julian et al., 2017; Auger et al., 2013; see also Epstein et al., 2014). While these studies examining PPA responses differ in their stimulus sets, designs, and even mode of presentation (e.g., pictures, mental imagery), they all leverage condition-rich designs and factor analyses to characterize the configuration of related object properties. While our observation of a role of spatial stability in PHC cannot in isolation lead to strong inferences, given that it is the product of a post-hoc analysis, it joins this collection of studies, adding to the diversity of stimulus sets and neuroimaging designs that show converging evidence that size in the context of position fixedness best explains the object responses in this region.

Our results, jointly with these collections of studies, suggest that a division based on manipulability for small objects and stability for large objects is fundamental in the organization of inanimate object responses.

4.2. What drives this object organization?

Turning to the broader question of *what drives* the spatial organization of object domains in visual cortex, as opposed to what visual properties are explicitly encoded in domain-specific regions, one answer is that the observed specialization is the result of evolutionary pressures

to maximize the efficient mapping of visual representations onto the appropriate downstream regions engaged in action relevant processing for a given object domain (Caramazza and Shelton, 1998; Conway, 2018; Konkle and Caramazza, 2017; Mahon et al., 2009; Mahon and Caramazza, 2011).

Candidate domains are manipulable (tools) and navigation-relevant objects. On this account, the representations computed in these regions are shape configurations and texture values that statistically reflect the distinguishing properties of manipulable versus navigation-relevant object domains. And, in the measure to which real-world size is differentially correlated with the two object domains (for example, large objects tend to be bulkier) it will have contributed to the evolving preferences for different visual features and configurations in the lateral and ventral occipitotemporal cortex. In other words, the two inanimate-object-prefering regions respond preferentially to, and presumably encode, the types of shape and texture properties that are associated with prototypical manipulable (small, graspable) objects and navigation-relevant (large, stable/stationary) objects.

It is relevant to note here that this view is fully compatible with accounts that emphasize the statistics of visual experience, wherein experienced eccentricity has played an evolutionary role in helping determine the localization of object domain preferences in the brain since it reflects the natural distribution of object viewing: more foveal, required for accurate reaching/grasping for small manipulable objects, versus more peripheral, associated with the requirement for capturing spatial (context) relations for large space-defining objects (Arcaro et al., 2009; Konkle and Oliva, 2012; Mahon and Caramazza, 2011; Malach et al., 2002).

A crucial distinction is being drawn here between, on the one hand, the factors that may have determined the observed neural specialization in different regions of OTC for different types of inanimate objects and, on the other hand, the information being computed/represented in those specialized areas. The role of the object property “size” at these two levels of description of neural specialization is illustrative in this regard. It could be argued that real-world object size, because of its role in distinguishing between manipulable and navigation-relevant object types, contributed to help select the visual shape/texture properties that characterize the observed domain specialization in OTC. But what about the role of the visual property “size” in characterizing the computations/representations in these neural regions? Is this visual property explicitly computed/represented in these areas? One reading of the available results is that size is not directly computed/represented in these areas.

There are at least three senses of the property size in the context of visual processing: subtended visual angle, perceived physical size, and real-world size. Konkle and Oliva (2012) found that only objects’ real-world size, and not retinal or inferred size, is related to regional specialization. However, this generalization is tempered by the observation, reported in Konkle and Caramazza (2013), that the effect of real-world size is limited to inanimate objects: real-world size differences in the animate domain are not associated with regional specialization. An implication of this observation, consistent with the results reported here, is that it is not real-world size, as such, that drives neural responses but the types of visual shapes that are typically associated with inanimate small, manipulable objects versus large, stable objects (e.g., Long et al., 2016). Furthermore, the property real-world size of an object is a “constructed” value that is retrieved from memory and has no obvious description in a vocabulary of visual properties such as shape, orientation, texture, and color. Indeed, the notion “real-world size” is a kind of conceptual/functional information like the properties “manipulability” and “navigation-relevance”, and none of these types of information is directly represented in the putative, size-responsive areas.

By positing a key role for domain-related processing, this view provides an account of data from congenitally blind individuals, who show response preferences in similarly located regions of occipitotemporal cortex when hearing the names of animate versus inanimate objects

or big versus small artifacts (e.g., Mahon et al., 2009; He et al., 2013; Peelen et al., 2013; Bi et al., 2015; Mattioni et al., 2020). Since ontogenetic visual experience could not have contributed to the similar patterns of object-prefering effects in high-level visual areas in congenitally blind and sighted individuals, it invites the conclusion that the observed specialization predates such experience. While the domain account provides a principled framework for explaining the potential functional basis for the observed large-scale organization of “visual” cortex, it is silent on the specific representational content at any given level of visual processing. Investigation of the specific perceptual features – e.g., elongated versus bulky shapes, texture properties, curvature patterns, or specific combinations of such features – that are represented/computed in the various object-type-prefering regions is an active, promising area of research (e.g., Andrews et al., 2010; Rajimehr et al., 2011; Bracci and Op de Beeck, 2016; Long, Yu and Konkle, 2018), aided by modeling with deep CNNs (e.g., Güçlü and Van Gerven, 2015; Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014; Kubilius et al., 2016).

Data and code availability statement

Data and Code are publicly available in the OSF open repository at: <http://doi.org/10.17605/OSF.IO/U8XFK>. All analyses run properly in Rstudio 1.3 and Matlab 2016b. Data includes: 1) preprocessed fMRI data, 2) fMRI activation maps (individual and group-level), and 3) object ratings from the Mechanical Turk Experiments.

Credit authorship contribution statement

Caterina Magri: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Talia Konkle:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Supervision. **Alfonso Caramazza:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

This work was supported in part by a grant from the Fondazione Cassa di Risparmio di Trento e Rovereto to the CIMeC – Centro Interdipartimentale Mente/Cervello, University of Trento (AC) and by the National Institutes of Health, National Eye Institute (Fellowship F32EY022863-01A1 to TK) and was conducted at the Laboratory for Functional Neuroimaging Center at the Center for Mind/Brain Sciences (CIMeC), University of Trento.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2021.118098](https://doi.org/10.1016/j.neuroimage.2021.118098).

References

- Andrews, T.J., Clarke, A., Pell, P., Hartley, T., 2010. Selectivity for low-level features of objects in the human ventral stream. *Neuroimage* 49 (1), 703–711.
- Arcaro, M. J., McMains, S. A., Singer, B. D., Kastner, S., 2009. Retinotopic organization of human ventral visual cortex. *Journal of neuroscience* 29 (34), 10638–10652.
- Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., ..., Penny, W., 2014. SPM12 manual. Wellcome Trust Centre for Neuroimaging, London, UK, p. 2464.
- Auger, S.D., Maguire, E.A., 2013. Assessing the mechanism of response in the retrosplenial cortex of good and poor navigators. *cortex* 49 (10), 2904–2913.
- Bainbridge, W.A., Oliva, A., 2015. Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage* 122, 408–416.
- Beauchamp, M.S., 2005. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15, 145–153.
- Bi, Y., Han, Z., Zhong, S., Ma, Y., Gong, G., Huang, R., ..., Caramazza, A., 2015. The white matter structural network underlying human tool use and tool understanding. *J Neurosci* 35 (17), 6822–6835.

- Boronat, C.B., Buxbaum, L.J., Coslett, H.B., Tang, K., Saffran, E.M., Kimberg, D.Y., De-
tre, J.A., 2005. Distinctions between manipulation and function knowledge of objects:
evidence from functional magnetic resonance imaging. *Cognitive Brain Research* 23
(2–3), 361–373.
- Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., Peelen, M.V., 2012. Closely
overlapping responses to tools and hands in left lateral occipitotemporal cortex. *J
Neurophysiol* 107 (5), 1443–1456.
- Bracci, S., Peelen, M.V., 2013. Body and object effectors: the organization of object repre-
sentations in high-level visual cortex reflects body–object interactions. *J Neurosci* 33
(46), 18247–18258.
- Bracci, S., de Beeck, H.O., 2016. Dissociations and associations between shape and cate-
gory representations in the two visual pathways. *J Neurosci* 36 (2), 432–444.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat Vis* 10 (4), 433–436.
- Campanella, F., D’Agostini, S., Skrap, M., Shallice, T., 2010. Naming manipulable objects:
anatomy of a category specific effect in left temporal tumours. *Neuropsychologia* 48
(6), 1583–1597.
- Brett, M., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using
the MarsBar toolbox for SPM 99. *Neuroimage* 16 (2), S497.
- Caramazza, A., Shelton, J.R., 1998. Domain-specific knowledge systems in the brain: The
animate-inanimate distinction. *J Cogn Neurosci* 10 (1), 1–34.
- Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal
cortex for perceiving and knowing about objects. *Nat Neurosci* 2 (10), 913–919.
- Chen, Q., Garcea, F.E., Jacobs, R.A., Mahon, B.Z., 2018. Abstract representations of objec-
t-directed action in the left inferior parietal lobule. *Cerebral Cortex* 28 (6), 2162–2174.
- Chen, Q., Garcea, F.E., Mahon, B.Z., 2016. The representation of object-directed action
and function knowledge in the human brain. *Cereb Cortex* 26 (4), 1609–1618.
- Conway, B.R., 2018. The organization and operation of inferior temporal cortex. *Annual
review of vision science* 4, 381–402.
- Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is
causally and selectively involved in scene perception. *J Neurosci* 33 (4), 1331–1336.
- Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for
visual processing of the human body. *Science* 293 (5539), 2470–2473.
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment.
Nature 392 (6676), 598–601.
- Epstein, R.A., Bar, M., Kveraga, K., 2014. Neural systems for visual scene recognition.
Scene Vision 105–134.
- Gallivan, J.P., McLean, D.A., Valyear, K.F., Culham, J.C., 2013. Decoding the neural me-
chanisms of human tool use. *Elife* 2, e00425.
- Garcea, F.E., Mahon, B.Z., 2014. Parcellation of left parietal tool representations by
functional connectivity. *Neuropsychologia* 60, 131–143.
- Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action.
Trends Neurosci 15 (1), 20–25.
- Grill-Spector, K., 2003. The neural basis of object perception. *Curr Opin Neurobiol* 13,
159–166.
- Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal
cortex and its role in categorization. *Nat Rev Neurosci* 15 (8), 536.
- Güçlü, U., van Gerven, M.A., 2015. Deep neural networks reveal a gradient in the com-
plexity of neural representations across the ventral stream. *J Neurosci* 35 (27),
10005–10014.
- He, C., Peelen, M.V., Han, Z., Lin, N., Caramazza, A., Bi, Y., 2013. Selectivity for large
nonmanipulable objects in scene-selective visual cortex does not require visual expe-
rience. *Neuroimage* 79, 1–9.
- Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. *Psychome-
trika* 30, 179–185.
- Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed rep-
resentation of objects in the human ventral visual pathway. *Proc Natl Acad Sci* 96
(16), 9379–9384.
- Julian, J.B., Ryan, J., Epstein, R.A., 2017. Coding of object size and object category in
human visual cortex. *Cereb Cortex* 27 (6), 3095–3109.
- Kalénine, S., Buxbaum, L.J., 2016. Thematic knowledge, artifact concepts, and the left
posterior temporal lobe: Where action and object semantics converge. *Cortex* 82,
164–178.
- Kellenbach, M.L., Brett, M., Patterson, K., 2003. Actions speak louder than functions: the
importance of manipulability and action in tool representation. *J Cogn Neurosci* 15
(1), 30–46.
- Khaligh-Razavi, S.M., Kriegeskorte, N., 2014. Deep supervised, but not unsupervised, mod-
els may explain IT cortical representation. *PLoS Comput Biol* 10 (11), e1003915.
- Konkle, T., Caramazza, A., 2013. Tripartite organization of the ventral stream by animacy
and object size. *J Neurosci* 33 (25), 10235–10242.
- Konkle, T., Caramazza, A., 2017. The large-scale organization of object-responsive cortex
is reflected in resting-state network architecture. *Cereb Cortex* 27 (10), 4933–4945.
- Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipi-
totemporal cortex. *Neuron* 74 (6), 1114–1122.
- Kubilius, J., Bracci, S., de Beeck, H.P.O., 2016. Deep neural networks as a computational
model for human shape sensitivity. *PLoS Comput Biol* 12 (4), e1004896.
- Lingnau, A., Downing, P.E., 2015. The lateral occipitotemporal cortex in action. *Trends
Cogn Sci* 19 (5), 268–277. doi:10.1016/j.tics.2015.03.006.
- Long, B., Konkle, T., Cohen, M.A., Alvarez, G.A., 2016. Mid-level perceptual features dis-
tinguish objects of different real-world sizes. *Journal of Experimental Psychology:
General* 145 (1), 95.
- Long, B., Yu, C.P., Konkle, T., 2018. Mid-level visual features underlie the high-level cate-
gorical organization of the ventral stream. *Proc Natl Acad Sci* 115 (38), E9015–E9024.
- Macdonald, S.N., Culham, J.C., 2015. Do human brain areas involved in visuomotor ac-
tions show a preference for real tools over visually similar non-tools? *Neuropsychologia*
77, 35–41.
- Mahon, B.Z., Anzellotti, S., Schwarzbach, J., Zampini, M., Caramazza, A., 2009. Cat-
egory-specific organization in the human brain does not require visual experience.
Neuron 63 (3), 397–405.
- Mahon, B.Z., Caramazza, A., 2011. What drives the organization of object knowledge in
the brain? *Trends in cognitive sciences* 15 (3), 97–103.
- Mahon, B.Z., Millelville, S.C., Negri, G.A., Rumiati, R.I., Caramazza, A., Martin, A., 2007.
Action-related properties shape object representations in the ventral stream. *Neuron*
55 (3), 507–520.
- Malach, R., Levy, I., Hasson, U., 2002. The topography of high-order human object areas.
Trends in cognitive sciences 6 (4), 176–184.
- Mattioni, S., Rezk, M., Battal, C., Bottini, R., Mendoza, K.E.C., Oosterhof, N.N., Col-
lignon, O., 2020. In: *Categorical representation from sound and sight in the ventral
occipito-temporal cortex of sighted and blind*, 9. Elife, p. e50732.
- Mruczek, R.E., von Loga, I.S., Kastner, S., 2013. The representation of tool and non-
tool object information in the human intraparietal sulcus. *J Neurophysiol* 109 (12),
2883–2896.
- Mullally, S.L., Maguire, E.A., 2011. A new role for the parahippocampal cortex in repre-
senting space. *J Neurosci* 31 (20), 7441–7449.
- Peelen, M.V., Bracci, S., Lu, X., He, C., Caramazza, A., Bi, Y., 2013. Tool selectivity in left
occipitotemporal cortex develops without vision. *J Cogn Neurosci* 25 (8), 1225–1234.
- Perini, F., Caramazza, A., Peelen, M.V., 2014. Left occipitotemporal cortex contributes to
the discrimination of tool-associated hand actions: fMRI and TMS evidence. *Frontiers
in Human Neuroscience* 8.
- Proklova, D., Kaiser, D., Peelen, M.V., 2016. Disentangling representations of object shape
and object category in human visual cortex: The animate–inanimate distinction. *J
Cogn Neurosci* 28 (5), 680–692.
- Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., Tootell, R.B., 2011. The “parahip-
pocampal place area” responds preferentially to high spatial frequencies in humans
and monkeys. *PLoS Biol* 9 (4), e1000608.
- Simmons, W.K., Martin, A., 2012. Spontaneous resting-state BOLD fluctuations reveal per-
sistent domain-specific neural networks. *Social Cognitive and Affective Neuroscience*
7 (4), 467–475.
- Striem-Amit, E., Vannuscorps, G., Caramazza, A., 2017. Sensorimotor-independent devel-
opment of hands and tools selectivity in the visual cortex. *Proc Natl Acad Sci* 114
(18), 4787–4792.
- Tarhan, L., Konkle, T., 2020. Reliability-based voxel selection. *Neuroimage* 207, 116350.
- Troiani, V., Stigliani, A., Smith, M.E., Epstein, R.A., 2014. Multiple object properties drive
scene-selective regions. *Cereb Cortex* 24 (4), 883–897.
- Wang, X., Peelen, M.V., Han, Z., He, C., Caramazza, A., Bi, Y., 2015. How visual is the
visual cortex? Comparing connectural and functional fingerprints between congeni-
tally blind and sighted individuals. *J Neurosci* 35 (36), 12545–12559.
- Whammoth, C., Chertkow, H., Murtha, S., Hanratty, K., 2002. Dissociable brain regions
process object meaning and object structure during picture naming. *Neuropsychologia*
40 (2), 174–186.
- Wickham, H., 2009. *ggplot2: elegant graphics for data analysis*. Springer New York 1 (2),
3.
- Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Perfor-
mance-optimized hierarchical models predict neural responses in higher visual cortex.
Proc Natl Acad Sci 111 (23), 8619–8624.