Symbolic categorization of novel multisensory stimuli in the human brain

Simone Viganò a,*, Valentina Borghesani b, Manuela Piazza a

a Centre for Mind/Brain Sciences, University of Trento, Italy
b Department of Psychology, Université de Montréal, Montréal, QC, Canada

A B S T R A C T

When primates (both human and non-human) learn to categorize simple visual or acoustic stimuli by means of non-verbal matching tasks, two types of changes occur in their brain: early sensory cortices increase the precision with which they encode sensory information, and parietal and lateral prefrontal cortices develop a categorical response to the stimuli. Contrary to non-human animals, however, our species mostly constructs categories using linguistic labels. Moreover, we naturally tend to define categories by means of multiple sensory features of the stimuli. Here we trained adult subjects to parse a novel audiovisual stimulus space into 4 orthogonal categories, by associating each category to a specific symbol. We then used multi-voxel pattern analysis (MVPA) to show that during a cross-format category repetition detection task three neural representational changes were detectable. First, visual and acoustic cortices increased both precision and selectivity to their preferred sensory feature, displaying increased sensory segregation. Second, a frontoparietal network developed a multisensory object-specific response. Third, the right hippocampus and, at least to some extent, the left angular gyrus, developed a shared representational code common to symbols and objects. In particular, the right hippocampus displayed the highest level of abstraction and generalization from a format to the other, and also predicted symbolic categorization performance outside the scanner. Taken together, these results indicate that when humans categorize multisensory objects by means of language the set of changes occurring in the brain only partially overlaps with that described by classical models of non-verbal unisensory categorization in primates.

1. Introduction

Animals have the ability to categorize objects into functional groups, and this is crucial to rapidly and efficiently select common behavioural schemes and to infer the correct response to similar but novel situations (Seger and Peterson 2013).

In the literature on the neural bases of categorization (based on intracranial electrophysiology in macaque monkeys (e.g. Rolls et al., 1977; Freedman et al., 2003; 2016) and non-invasive human neuroimaging (Op de Beeck et al. 2006; Jiang et al., 2007; Folstein et al., 2013; Brants et al., 2016)) there is a consensus that learning to categorize simple visual or acoustic stimuli by means of non-verbal matching tasks generates two types of changes in their brain. The first consists of an increased precision with which stimuli are encoded in sensory areas (infratemporal and lateral occipital cortices for visual stimuli and Heschl’s gyrus, superior temporal gyrus, and planum temporale for auditory stimuli (Ley et al., 2012; Jiang et al., 2018)). The second type of change is detected in parietal and prefrontal cortices and takes the form of a categorical response to stimuli: objects that belong to the same group are represented more similarly compared to those that belong to different ones, irrespective of their perceptual similarities. Interestingly, while the first type of change appears to be task-independent, the emergence of categorical responses is task-dependent, in that it is only present when the task explicitly requires categorization, and it is absent otherwise (Jiang et al., 2007; Jiang et al., 2018; Roy et al., 2010; van der Linden et al. 2014). The analysis of temporal latencies of time-sensitive measures further revealed that while selectivity to sensory features emerge between 120 and 200 msec from the stimulus onset, categorical effects are detectable later, about 170–200 ms after stimulus onset (Freedman et al., 2003; Scholl et al. 2014), consistent with the idea that categorization involves at least two stages, distinguishable both in terms of brain localization and temporal activation: an early stage of perceptual discrimination in early sensory cortices, followed by a categorical stage involved in categorical decision-making that takes place in parietal and prefrontal cortices (Freedman et al., 2003; Riesenhuber and Poggio 2000). The present study develops from this body of work and extends it to stimuli and tasks that are particularly relevant for the human species and that were not studied before: the case of multisensory stimuli and symbolic categorization.

With respect to the stimuli, previous studies focused on single sensory modalities, either vision or audition. However, in the natural environment objects typically present a multitude of features, and their multisensory nature permeates our experience: since childhood, we learn that the category “cat” refers to entities defined by the conjunction of a variety of multisensory features: they meow, have a soft fur, four legs, etc. Being able to integrate different sensory features improves our ability to categorize and later remember objects in the world (e.g., a picture of a cow is better recognized and remembered if it is combined with a sound “moo”, compared to when it’s presented without, or with inconsistent, sounds; for reviews see Shams and Seitz 2008; Matusz et al., 2017). However, to date the neurocognitive mechanisms of multisensory categorization have remained largely undescribed. The classical two-stages model of categorization (i.e., perceptual identification followed...
by categorical decision-making) does not make explicit predictions: is multisensory integration an extra stage of the categorization process? If so, where and how in the brain is this process resolved?

The second aspect that characterizes previous studies is the task used to assess categorization: non-verbal match-to-category tasks where subjects (monkeys or humans) are required to decide whether two stimuli belong to the same category or not (e.g., Freedman and Assad 2016). Although these procedures allow to isolate the access to categorical representations from the preparation/execution of a motor response, they fail to tackle a key aspect of human categorization, that is the pervasive use of symbolic labeling (naming). In ecological settings, humans mainly categorize the world using language. When seeing a cat and a dog crossing the street one refers to the experience stating: “I have seen a cat and a dog”, and not: “I have seen two animals that are not of the same category”. Thus, not taking into account the role played by symbols in categorization results in a partial description of how categorization works in our species. From a behavioural point of view, we know that symbols facilitate categorical learning (Xu 2002; Waxman and Hall 1993, Waxman and Markow 1995, Althaus & Westermann 2016; Yamauchi and Markman 2000; Lupyanc et al., 2007), abstraction (Edmiston and Lupyanc 2015), and object recognition (Lupyanc and Ward 2013; Boutonne and Lupyanc 2015). However, the neural correlates and consequences of using symbols during categorization are largely unexplored. One possibility is that learning to categorize through symbols generates a common shared neural representation between objects and their corresponding symbolic form (e.g. see the case of specific quantities - sets of objects, and their symbolic form - Arabic digits, e.g., Piazza et al., 2007; Eger et al., 2009). This stage should be integrated in a revised extended model of categorization.

We hypothesized that categorization of multisensory stimuli through symbols entails at least three representational stages: (i) perceptual discrimination of unique sensory features (e.g., audio vs video); (ii) multisensory integration of sensory features into unique object identities (audio + video); (iii) association of objects classes to their categorical name (audiovisual object + label).

After a behavioural training aimed at teaching participants how to classify novel audiovisual objects into four labelled categories, we scanned them with functional Magnetic Resonance Imaging (fMRI). We presented them with the objects and the words, in pseudo-random sequence, engaging them in a cross-format category repetition task that required to detect anytime an object was preceded or followed by its categorical name. This task guaranteed the access to symbolic categorical identity for each stimulus, irrespective of the actual need to give a response. An fMRI recording session also took place before learning, when participants were presented with the same audiovisual objects and categorical names without knowing their mutual correspondence. During this pre-learning session, participants performed a simple 1-back stimulus repetition task, requiring detecting anytime two following stimuli were identical. The purpose of this session was to collect the brain response to objects and words before they acquired any categorical meaning thus to serve as a comparison/control for results obtained during the post-learning symbolic categorization task, main objective of the study.

2. Methods

2.1. Participants

We tested 25 right-handed adult volunteers (fifteen females; mean age = 22.20, std = 2.74). They gave written informed consent, underwent screening to exclude incompatibilities with the MRI scanner, and were reimbursed for their time. The study was approved by the ethics committee of the University of Trento (Italy). Data from 4 subjects were excluded from the analyses given their poor behavioural performance during the second fMRI day (accuracy < 70%). This led to a final sample of 21 participants (thirteen females; mean age = 21.95, std = 2.58) which, assuming an alpha-value of 0.05 for two-tailed one-sample t-test and a power of 0.8, led to an effect size of 0.64 (calculated with G’Power).

2.2. Stimulus space

We developed a set of 16 novel animated multisensory objects, orthogonally manipulating the size of an abstract shape (Fig. 1A) and the pitch of an associated sound. While in the natural world different sensory features tend to vary in a correlated manner (e.g. big animals are usually slower, or they produce low-frequency sounds), here we made the two sensory features (size and pitch) vary orthogonally because this allowed to empirically test how the information from different sensory systems is represented in the brain. A total of four size- and pitch-levels were used for each participant, leading to a stimulus space where each object represented the unique combination of one size and one pitch level (Fig. 1B). The values of these two features were selected for each participant on the first day of the experiment, following a brief psychophysical validation consisting of a QUEST adaptive staircase method (Watson & Pelli 1987)(Supplementary Figure S1). The rationale of this first validation was to equalize subjects’ perception of the perceptual features. Using a two-stimuli comparison task for size and for pitch, we calculated subject-specific sensitivity as the Just Noticeable Difference (JND) from a reference value (size: visual angle of 5.73°, pitch: frequency of 800 Hz) leading to 80% of correct responses. For each sensory feature, four subject-specific feature levels were calculated, applying Weber-Fechner’s law and selecting values at every three JNDS, to ensure that feature levels were equally clearly discriminable. In order to strengthen the multisensory binding between the two unisensory features, we applied a ‘squeezing’ animation during each object presentation by displaying 13 frames of the same object with increasing (frames 1 to 7) and decreasing (frames 8 to 13) size along the horizontal axis. This conveyed the impression that the sound was produced by the animated object. The objects were presented for 750 ms and the sounds were presented at the apex of the squeezing period. The objects were divided into four categories based on the combination of two categorical boundaries (Fig. 1B). The categorical membership of each object, as well as their unique multisensory identities, could thus be recovered only when considering the two sensory features in combination. We assigned to each category a name (three-letters pseudoword) (Fig. 1C): KER (small size and low pitch); MOS (big size and low pitch); DUN (small size and high pitch); GAL (big size and high pitch). The bidimensional arrangement of objects was never made explicit to subjects.

2.3. Stimuli presentation

Stimuli were presented foveally using MATLAB Psychtoolbox in all experimental phases, at a distance of ~ 130 cm. Multisensory objects’ size and pitch ranged from 5.73° and 800 Hz for level 1 to an average of 8.97° and 973.43 Hz for level 4. Each word subtended a visual angle of 3.58° horizontally and 2.15° vertically, and was presented with black Helvetica font on a gray background.

2.4. Procedure

2.4.1. Experimental sessions

The experiment consisted of three parts: pre-learning fMRI, behavioural training, and post-learning fMRI (Fig. 1D). During pre-learning fMRI, participants were exposed for the first time to the new multisensory objects and to the novel names. Starting with the following day, subjects underwent nine sessions of behavioural training outside the scanner. The aim was to teach them the object-name correspondence, an operation requiring parsing of the object space into four categories and connecting each symbol (word) to its meaning (the corresponding category exemplars). Finally, during the post-learning fMRI, subjects were again exposed to the same objects and words. On average, the second fMRI session occurred 9.86 days (std = 1.4 days) after the first one. All
the tasks are described below. During both fMRI sessions, and during the first 8 training days, we used 8 out of 16 objects available in each subject’s stimulus space (objects: 1 - 3 - 6 - 8 - 9 - 11 - 14 - 16; see Fig. 1B); the remaining 8 were used only during the 9th training session to test for generalization of the categorical rule to novel exemplars.

2.4.2. Functional localizer

At the start of the pre-learning fMRI session, participants underwent a block-design functional localizer designed to isolate the cortical regions recruited to process visual and acoustic components of our objects. During video mini-blocks, participants were presented with animated objects varying in their size, without any acoustic component. During audio mini-blocks participants were presented with sounds varying in pitch, without the object visual component. There were four blocks for each condition (video or audio), resulting in a total of 8 mini-blocks of six stimuli each, presented in pseudo-random order. Each block was preceded and followed by 10 s of fixation cross. For each block, participants had to perform a simple 1-back task, pressing a button whenever they detected a repetition of the same stimulus: same size for video-blocks, same pitch for audio-blocks.

2.4.3. fMRI tasks

During the first fMRI session, before training, participants performed an oddball stimulus repetition detection task, where they were presented with the multisensory objects and the novel words in pseudorandom order. They were instructed to press a button when they detected a rare immediate repetition of the very same stimulus (either a multisensory object or a word). Each stimulus was presented for either 750 ms (objects) or 500 ms (words), with a variable ISI of 4 ± 1.5 s during which a blue fixation cross was presented. There were 4 runs, each one lasting around 7 min. Within a run, each stimulus was repeated 6 times, resulting in 72 trials per run. There was one target event (1-back repetition) per stimulus, for a total of 12 out of 72 (17%) stimulus repetition per run. During the second fMRI session, after training, participants were presented again with the same stimuli, now performing a symbolic categorization task, where they had to detect a rare immediate repetition of the same category when expressed as a word and as object (e.g. object 1 followed by the word “KER”). This task, requiring the explicit recovery of the learnt object-name association, simulated the training task, and could not be performed before learning given the absence of any association between words and object categories. This resulted in a total of 16 target events (~22%) per run. The number of runs, trials, and stimuli repetition for this task were identical to those presented during the first fMRI day.

2.4.4. Behavioral training

Participants underwent 9 daily sessions of behavioral training. Each behavioral session lasted 10 min, and it was divided into 4 mini-blocks of 20 trials each. It started with a brief presentation of the objects as exemplars of the four categories: a sentence like “These are exemplars of the category KER” was visually presented and was followed by the two objects exemplars, presented in sequence, coming from the same category. After this familiarization phase, each trial consisted of an object presentation (750 ms), followed by a fixation cross (500 ms), and by the presentation of the 4 possible names in random order, from left to right (Fig. 1F). Each object was presented 10 times per training session. Participants were instructed to press one key on the keyboard to select the correct name. They were asked to respond as fast as possible, but no time limits were imposed. After their response, an immediate feedback appeared on the screen for 1000 ms, indicating with the
words “Correct!” or “Wrong!” the accuracy of the choice. In the case of a wrong answer, the feedback also showed the correct object name. After each miniblock, participants would be provided with the cumulative percentage accuracy. Starting from the seventh training session, the trial-by-trial feedback was removed. For the first 8 days of training, participants were presented with the same 8 objects used in the two fMRI sessions. On the last training day, without being notified of the change, they were presented with all 16 objects. This allowed us to test for generalization of the categorical rule to new exemplars (here represented by objects 2 · 4 · 5 · 7 · 10 · 12 · 13 · 15), which is a key ingredient of an efficient semantic representation. For this last session, the mini-blocks number was kept at 4, but the number of trials was doubled.

2.5. Neuroimaging acquisition

Data were collected on a 4T Bruker scanner (Bruker BioSpin) with standard head coil, at the Center for Mind/Brain Sciences, University of Trento, Italy. Functional images were acquired using EPI T2-weighted scans. Acquisition parameters were as follows: TR = 3 s; TE = 21 ms; FA = 81°; FOV = 100 mm; matrix size = 64 × 64; number of slices per volume = 51, acquired in interleaved ascending order; voxel size = 3 × 3 × 2 mm. T1-weighted anatomical images were acquired twice per participant (pre- and post-learning) with an MP-RAGE sequence, with 1 × 1 × 1 mm resolution.

2.6. Preprocessing and general linear model

Functional images were preprocessed using SPM8 in MATLAB. Preprocessing included realignment of each scan to the first of each run, co-registration of functional and session-specific anatomical images, segmentation, and normalization to the MNI space, as implemented in SPM8. No smoothing was applied. Functional images for each participant individually were analyzed using a general linear model (GLM) separately for the two fMRI sessions and for the different tasks. For each run we included 13 regressors of interest, corresponding to the onsets of the eight objects, the four words, and the motor response; 6 regressor for head-movements (estimated during motion correction in the preprocessing); 3 regressors of no interest (constant, linear, and quadratic trends). No pair of regressors had a correlation higher than 0.14. Baseline periods were modelled implicitly, and regressors were convolved with the standard HRF without derivatives. A high-pass filter with a cutoff of 128 s was applied to remove low-frequency drift. We thus obtained one beta map for each stimulus (eight objects and four words) for each run. The following multivariate analyses were performed using the CoSMoMVPa toolbox (Oosterhof et al. 2016) for Matlab.

2.7. Regions of interest

Regions of interest (ROIs) responding to the visual and to the acoustic components of the multisensory objects were isolated on the pre-learning imaging data and were used for our first analysis (see below). We selected brain activity evoked at group-level (p < .001, FWE corrected at q < .05) by objects presentation during the 1-back task on stimulus identity (pre-learning task), masked with the group-level results (p < .001, FWE corrected at q < .05) of the functional localizer for either the visual and the acoustic modality. This resulted in a bilateral network wherein the LateralOccipitalComplex (LOC) and the anterior portions of the SuperiorTemporalGyrus (STG) responded to the visual and to the acoustic components of our stimuli, respectively (Fig. 2A). All clusters were binarized and used as regions of interest in the following analyses.

2.8. Sensitivity to sensory features in sensory ROIs

Our first question was whether the sensory regions that responded preferentially to the visual and auditory components of the stimuli before learning (see above), during post-learning symbolic categorization task (i) displayed some form of multisensory integration (e.g. visual areas encoding size also representing differences in pitch) and/or (ii) displayed enhanced sensitivity, compared to pre-learning, to their preferred sensory modality (e.g., visual areas better representing differences along size), to the non preferred one, or to both. We first extracted, using CoSMoMVPa (Oosterhof et al. 2016), the neural dissimilarity (1 - Pearson’s correlation) between pairs of objects varying along one sensory dimension only (e.g. between object 1 and object 3, that varied in their size but had the same associated sound, Fig. 2B) from the post-learning symbolic categorization task. To answer to the first question we compared, for each ROI, the neural dissimilarity between objects varying along size vs those varying along pitch. To answer the second question we compared the post- to the pre-training data. We computed the difference in representational dissimilarity for objects varying along size or pitch between the two fMRI sessions.

2.9. Multisensory integration of sensory features into individual object representations

Our second question was to isolate the brain regions that developed a representation of the unique identities of the 8 different audiovisual objects. To do so we applied a multivariate approach (Haxby et al. 2014), implemented using a whole brain searchlight. A sphere with a radius of 3 voxels - selected for consistency with previous studies (Connolly et al. 2012, Fairhall & Caramazza 2013) - was centered in every voxel of the subject- and session-specific datasets. Within each sphere, we conducted a split-half correlation analysis (Haxby et al. 2001) that allows for testing whether the distributed activity differentiates between stimuli: we extracted the patterns of BOLD activity evoked by the presentation of the 8 objects separately. Then we divided the dataset into two halves, and we crossed the neural representations of each object with each other, resulting in a correlation matrix with 8 × 8 entries (Fig. 3A); here, the correlation between matching stimuli coming from the two different halves of the fMRI data laid on-diagonal, while the correlation between non matching stimuli laid off-diagonal. If the brain activity in any given brain region represents each object as unique identity, then the mean difference between Fisher-transformed values on-diagonal versus off-diagonal, resulting from all the possible combinations of the two dataset halves, should be positive. For each sphere of the searchlight, the resulting correlation score was stored in the center voxel, thus leading to one correlation map per subject and per session. Single-subjects’ correlation maps were then submitted to a group-level analysis, as implemented in the “second-level analysis” option of SPM8, thresholded at p < .001, FWE corrected at q < .05. To be sure that the resulting clusters were genuinely sensitive to multisensory information and not solely to one or the other sensory feature alone (that is, differentiating objects only based on their size or on their pitch), we additionally run two corresponding searchlights searching for brain regions responding to unimodal variations between objects: more specifically, we looked for brain regions showing high similarity between different objects that shared the same size or the same pitch level, respectively. Same-object similarity (on diagonal) was excluded because it did not allow us to distinguish between similarity in size or pitch alone. These two resulting maps (see Supplementary Figure 3A-B) were merged and used as a single exclusive mask for the group-level analysis on object identities, therefore guaranteeing that the resulting clusters were genuinely sensitive to conjunctive multisensory representations. Fig. 3C offers a confirmatory visualization of this control in the brain regions where we detected the representation of individual object identities. For the following analyses, we created spherical ROIs with a radius of 3 voxels around the peaks of the group-level analysis, to ensure matching voxel size of our ROIs. Corresponding results were obtained when the entire clusters were considered.
Fig. 2. A. Sensory ROIs. The visual and the acoustic components of the objects before learning were separately processed in lateral occipital cortex (LOC) and anterior superior temporal gyrus (aSTG), respectively, as revealed by a functional localizer (see Methods). These areas were binarized and used as regions of interest for subsequent analyses. B. Comparisons along single sensory dimensions. Neural dissimilarity (1-Pearson’s r) was computed between those pairs of objects that varied along one sensory modality solely, during the post learning symbolic categorization task, such as objects 1 and 3, that varied only along size, or objects 6 and 14, that varied only along pitch. C. Sensory specialization. The representational dissimilarity between objects varying along size was greater in LOC, while that between objects varying along pitch was greater in aSTG. ∗∗∗ p < 0.005. D. Sensory segregation. When compared between post and pre learning, the neural dissimilarity between objects varying along size was higher in LOC, while that between objects varying along pitch was higher in aSTG. In particular, for a given ROI, the neural dissimilarity between objects varying along its preferred sensory modality increased after learning, while it decreased for objects varying along the non preferred modality. ∗ < 0.05.

Fig. 3. A. Split-half correlation analysis for object identities. Using a combination of multivariate split-half correlation analysis and whole-brain searchlight (see Methods) we revealed three brain regions where individual object identities were represented as unique audiovisual combinations during the post-learning symbolic categorization task: left angular gyrus (AG), left middle frontal gyrus (MFG), and right inferior frontal gyrus (IFG). B. Multisensory object identities emerged only after learning. None of these regions represented individual object identities before learning, as revealed by the same multivariate split-half correlation analysis applied to the activity of these three ROIs on the pre-learning fMRI data. C. Object identities were not based on unisensory details - These areas did not represent more similarly those objects that had the same size (top matrix) or the same pitch (bottom matrix).
2.10. Functional connectivity with beta-series correlation

To explore the possibility that multisensory integration emerged from the functional connectivity between sensory areas, we implemented a beta-series correlation approach, optimized for event-related designs (Rissman, Gazzaley, D’Esposito 2004; Cisler, Bush, Steele 2014). We run a separate GLM modeling each trial separately, resulting in a single beta-series for each one of them. We averaged the beta-series across voxels within each sensory ROI and correlated them to quantify functional connectivity.

2.11. Association of each object to its category name

Next we asked whether and where in the brain the symbolic category training and task produced a common representational code between objects and their category name. We started with an ROI approach looking at the brain regions where multisensory object identities were detected. We extracted and correlated the distributed responses to the words and the object categories, under the assumption that the activity evoked by a given set of objects (e.g., objects 1 and 3, belonging to the category “KER”) should be more similar to the one evoked by the matching category name (e.g., the word “KER”), compared to all other words. Thus, for each subject, we stored the mean difference between Fisher-r-to-z-transformed on-diagonal vs. off-diagonal values of this 4 × 4 matrix (see Fig. 4A) and tested the data against the null hypothesis of no difference in correlation at the group level. Additionally, to avoid overlooking other potential brain regions that could contribute to this abstract categorical representation, we implemented the same analysis in a whole brain searchlight (for parameters, see above Split-half correlation analysis).

2.12. Assessing abstraction: cross-format decoding (words to objects and objects to words) and decoding stimulus format (words vs. objects)

To assess the level of abstraction of the common representational code, we implemented two additional analyses. First, we trained a Linear Discriminant Analysis (LDA) classifier, as implemented in the CoSMoMVAs toolbox (Oosterhof et al. 2016) to distinguish between categories in one format (e.g., objects, or words) and we tested it on the other format (e.g., words, or objects). This would indicate that the shared representation between matching objects and names is enough to generalize categorical knowledge across presentation modalities. We stored the average accuracy across formats for each subject for the following group-level test against a null hypothesis of chance classification (4 categories, 25%). Next, we quantified the residual information about the presentation format, indicated by the accuracy of a classifier to decode whether, at any given trial, subjects were presented with a word or an object.
We stored each subject’s and ROI’s accuracy for a later group-level test against a null hypothesis of chance performance (50%)."

2.13. Representation similarity of within-category objects

To quantify whether and where two objects belonging to the same category were represented more similarly to each other compared to those belonging to different groups, we used model-based Representational Similarity Analysis (RSA) (Kriegeskorte et al. 2008). We computed for each pair of objects their dissimilarity (1 - Pearson’s correlation) resulting in a 8 x 8 matrix with meaningless diagonal. Then, we correlated the Fisher-r-to-z-transformed lower triangle of this neural dissimilarity matrix (neural DSM) to the lower triangle of a 8 x 8 model assuming maximum dissimilarity (value 1) between all pairwise objects except for those 4 pairs representing members of the same category, for which we assumed minimum dissimilarity (value 0) (Fig. 4D). We implemented this analysis both within the ROIs emerging from the previous steps, and within a whole brain searchlight with matching parameters as the ones described above. For the group level whole brain analysis, we used the same exclusive mask that we used for the split-half correlation analysis for object identities: this was motivated by the fact that the objects within categories share perceptual features that could bias the results.

3. Results

3.1. Behavioural results

The behavioural training consisted of 9 daily sessions outside the scanner, lasting ~20 min each (see Methods), during which participants performed a simple delayed-match-to-category-name task (Fig. 1F, see Methods). The learning trajectory indicates an increment in performance from session 1 to 4, while from session 5 on participants maintained unchanged their accuracy level (session 1: 61 ± 18%; session 2: 76 ± 17%; session 3: 83 ± 13%; session 4: 87 ± 11%; session 5: 89 ± 10%; session 6: 90 ± 11%; session 7: 91 ± 8%; session 8: 91 ± 7%), indicating a period of consolidation (Supplementary Figure 1D). Performance collected on the last behavioural session confirmed both the successful learning of the name-objects association and its generalization. Performance was high for both sets (training set: mean accuracy = 88.12%, std = 5.93%, different from chance t(20) = 29.44, p < .001; generalization set: mean accuracy = 75.03%, std = 10.68%, difference from chance, t(20) = 10.74, p < .001), even though it was lower for the generalization set (t(20) = 5.06, p < .001) (Supplementary Figure 1D). Behavioural performance during the fMRI tasks was as follows: during pre learning fMRI session (1-back stimulus identity task), mean accuracy was 92.96% (std = 5%), mean reaction time was 521 ms (std = 165 ms); during the post-learning fMRI session (symbolic categorization task) mean accuracy was 86.98% (std = 8%), mean reaction time was 714 ms (std = 222 ms). Both accuracy and RTs significantly differed across the two tasks (accuracy: t(20) = 3.26, p = .004; RTs: t(20) = 4.75, p < .001).

3.2. Neuroimaging results

3.2.1. Sharpening and segregation of sensory features (Stage 1)

Motivated by previous studies (Op de Beeck et al. 2006; Jiang et al., 2007; Folstein et al., 2013; Brants et al., 2016; Ley et al., 2012; Jiang et al., 2018), we asked whether in the case of multisensory objects visual and acoustic cortices sharpened neural selectivity of the visual and acoustic components of the objects. Whole-brain activation during objects’ presentation (computed by the contrast objects vs. baseline) were almost identical in the two sessions (see Supplementary Figure 2A-B): a direct comparison between post-learning and pre-learning session revealed that no brain region was significantly more active after learning compared to pre learning or vice versa (p < .001, FWE corrected at cluster level with q < 0.05). Thus, we approached our questions with a combination of ROI-based and multivariate pattern analysis. First we selected, through an independent functional localizer (see Methods), visual and acoustic ROIs - bilateral Occipital Cortex (LOC) and anterior Superior Temporal Gyrus (STG) respectively - responding selectively to the visual and acoustic components of the multisensory objects before learning took place (see Methods) (Fig. 2A). Next, in each of the two sensory ROIs, we compared the distributed patterns of activity evoked by objects that varied along the two sensory modalities separately (e.g. object 1 and 3 varied only along size, while objects 1 and 9 only along pitch), during the post-learning cross-modal category repetition task (Fig. 2B). We computed the representational dissimilarity (1 - Pearson’s r) (see Methods) between all these object pairs, to quantify how sensitive these regions were to differences along either size or pitch. Results revealed that differences along size were represented more strongly in LOC than in aSTG (size in LOC: t(20) = 7.40, p < .001; size in aSTG: t(20) = 2.11, p = .047; t-tests, alpha-corrected (Bonferroni) = 0.05/2 = 0.025; LOC vs aSTG: t(20) = 3.05, p = .006; t-test) while differences along pitch were represented more strongly in aSTG than in LOC (pitch in aSTG: t(20) = 7.68, p < .001; pitch in LOC: t(20) = -0.25, p = .80, alpha-corrected (Bonferroni) = 0.05/2 = 0.025; aSTG vs LOC: t(20) = 4.55, p < .001; t-test) (Fig. 2C). Next, we asked whether and how the selectivity for each feature category changed in the two ROIs compared to pre-learning. When considering the post-pre representational dissimilarity between objects, there was a significant interaction between sensory feature (dissimilarity along size vs along pitch) and ROI (LOC vs aSTG) (Fig. 2D). Indeed, consistently with previous reports, the representational distance between objects varying along size increased in size-sensitive regions (LOC; post-pre: t(20) = 1.81, p = .043, t-test), while the representational distance between objects varying along pitch increased in pitch-sensitive regions (aSTG; post-pre: t(20) = 1.82, p = .042, t-test) (Fig. 2D). Concurrently however, we observed a previously overlooked set of results: the representational distance between objects varying along size decreased in aSTG (t(20) = -1.6, p = .063, t-test), and that between objects varying along pitch decreased in LOC (t(20) = -2.15, p = .02, t-test). The post-pre sensitivity to size differences was significantly different between LOC and aSTG (t(20) = 2.39, p = .02, t-test), as well as their sensitivity in pitch (t(20) = 2.46, p = .023, t-test) (Fig. 2D). This indicates that training resulted in an increased sensory segregation in sensory areas. These results were not associated with differences across scanning sessions in the temporal Signal-to-Noise Ratio (tSNR) (pre- vs post-learning) in none of the ROIs (LOC vs aSTG) (LOC pre-learning: average = 39.79 (std = 5.49); LOC post-learning: average = 39.84 (std = 4.38), difference post ≠ pre: t(20) = 0.63, p = 0.53; aSTG pre-learning: average = 39.81 (std = 4.24); aSTG post-learning: average = 40.09 (std = 4.04), difference post ≠ pre: t(20) = 1.62, p = 0.12).

3.2.2. Multisensory integration of sensory features into unique audiovisual object identities (Stage 2)

We then asked whether and where in the brain the audiovisual features were integrated to form unique object identities. A split-half whole brain searchlight correlation analysis (see Methods) during the symbolic categorization post-learning task revealed that three brain regions developed a local and distributed code for each unique individual object identities: the left angular gyrus (MNIx,y,z = -36, -67, 26; t(20) = 6.8, cluster size (k) = 1434), the left middle frontal gyrus (MFG)(MNIx,y,z = -33, 47, 18; t(20) = 7.82, k = 2288), and the right inferior frontal gyrus (IFG)(MNIx,y,z = 51, 35, -6; t(20) = 6.98, k = 1535; all p < .001, FWE corrected at q < 0.05) (Fig. 3A). Interestingly, these regions did not show the same effect before learning (Fig. 3B). These regions did not represent similarities among individual sensory dimensions (Fig. 3C). Unexpectedly, no other brain region showed evidence of local multisensory integration before learning, as revealed by a whole brain searchlight. This was in apparent contradiction with the high behavioural performance of subjects, who were able to detect object identities before learning at high accuracy (92.96% (std = 5%)). One possibility is that, before learning, participants might have processed acoustic and visual information
separately, i.e. via the concurrent activation of segregated unisensory representation. This would potentially predict a stronger functional connectivity between sensory regions encoding size and pitch before compared to after training. A beta-series correlation approach confirmed the hypothesis (pre-learning: t(20)=2.9, p=.008; post learning, t(20) = 0.9, p = .37; pre>post: t(20) = 1.73, p = .048, t-tests)(see Methods).

3.2.3. A shared representational code common to objects and their categorical names (Stage 3)

Next, we looked for signatures of the association between object categories and their specific names, the last essential stage of symbolic categorization. To test for this, we implemented a cross-format correlation analysis (see Methods), looking for brain regions where the activity patterns evoked by each (category of) objects were more similar to those evoked by the corresponding category words compared to the other words (Fig. 4A). First, we applied the analysis in the three ROIs encoding object identities. Only the left angular gyrus displayed this category-specific, format invariant coding property, indicating that this area responded similarly to objects and the corresponding words, thus encoded their learnt association (left AG: t(20) = 3.53, p = .002; MFG: t(20) = 0.9, p = .36; IFG: t(20) = 1.73, p = .10; t-tests, alpha-corrected(Bonferroni) = 0.05/3 = 0.016). This common response for objects and words in the left angular gyrus was absent before learning (t(20) = 0.56, p = .58; t-test).

To take a step further, we investigated whether any brain region developed a similar common representation between symbols and their referents. A whole-brain searchlight confirmed the ROI results (left angular gyrus, also extending to the superior parietal lobule (MNIx,y,z: −27, −58, 46; t(20) = 6.6; k = 1056) and extended them to the right hippocampus (MNIx,y,z: 27, −22, −6; t(20) = 5.14, k = 68) (Fig. 4A). These two regions have been previously involved in conceptual/semantic processing and relational memory (e.g., Binder et al., 2009; Fairhall & Caramazza 2013; Zeithamova et al., 2008; Mack et al., 2016; Zeithamova et al., 2012; Theves et al., 2020). This signal was also observed in an additional network of brain regions comprising the left fusiform gyrus/ventral occipitotemporal (vOT) (MNIx,y,z: −51, −61, −12; t(20) = 6.7, k = 187), the right superior and inferior parietal lobule (MNIx,y,z; 21, −73, 54; t(20) = 6.6, k = 170; and MNIx,y,z: 57, −46, 46; t(20) = 5.1, k = 62; respectively), and the right angular gyrus (MNIx,y,z: 54, −67, 34; t(20) = 5.25, k = 95; all p < .001, FWE corrected at q<0.05) (Supplementary Figure 4). Before learning, none of these regions showed this kind of format-invariant coding of object category (all p > .18). Supplementary Figure 5 shows that for each object, the evoked pattern activity was more similar to the one evoked by the matching name, compared to the remaining non-matching ones, thus showing that this effect was not triggered by few object exemplars solely.

3.2.4. Levels of abstraction in the representational format of the categories

Next, we wanted to better characterize the levels of abstraction in representing object-name associations during symbolic categorization. In particular, of those displaying object-name associations, we focused on two brain regions specifically, the left angular gyrus and the right hippocampus (Fig. 4A), motivated by their known involvement in processing conceptual knowledge and relational memory. To do so, we applied two different decoding approaches. First of all, we verified whether the categorical information generalized across the two formats. We trained an LDA classifier to distinguish categories using the brain response to stimuli in one format (objects or words) and we tested it on the one evoked by the stimuli in the other format (words or objects). The average generalization accuracy in the right hippocampus (29.91%, std = 9.3%) was significantly above chance (t-test against chance level of 25%; t(20) = 2.42, p = .02)(Fig. 4B), and this was not the case before learning (accuracy: 22.47%, std = 7.9%, t-test against chance level of 25%; t(20) = −1.45, p = .16), with a significant difference between pre and post (t(20) = 2.76, p = .01, t-test). In the left angular gyrus, instead, generalization was not significant (average accuracy: 26.93%, std = 7.94%, t-test against chance level of 25%; t(20) = 1.11, p = .27)(Fig. 4B). However, there was no statistical difference in the generalization accuracy between the two ROIs (t(20) = 1.23, p = .23, paired t-test).

Next, we quantified the residual information about stimulus format. This was a relevant piece of information to describe the degree of sensitivity to perceptual details of a stimulus in each region: indeed, efficient abstraction might benefit from a lower sensitivity to perceptual details, such as the format of the incoming stimulus, to enhance its abstract, or conceptual, content. To this end we trained the classifier to predict the format of the incoming stimulus (object or words). The classifier accuracy was very high in the left angular gyrus (mean accuracy = 59%, t(20) = 5.15, p < .001, t-test) but it did not significantly diverge from chance in the right hippocampus (mean accuracy = 52%, p = .12, t-test)(Fig. 4C). Here the difference between the two regions was significant (L-AG vs R-HPC: t(20) = 3.22, p = .004; paired t-test). As a final step, we reasoned that if the degree of category-specific correspondence across stimulus formats is functionally relevant, it should predict the level of behavioural accuracy achieved after training in detecting cross-format category correspondences (those subjects who performed best in the object-category name association task should be those ones which neural similarity across format is higher). We tested this prediction by correlating the individual subjects’ neural object-name similarity score with their behavioural performance at the end of training (test day), when, contrary to the task they performed in the scanner, they were asked to give a response at every trial, thus allowing us to well estimate their learning. Results showed a positive correlation in the hippocampus (r = 0.58, p = .006) but not in the left angular gyrus (r = 0.14, p = .54)(Fig. 4D). None of the other ROIs we found with the cross-modal correlation analysis showed significant correlation with behavior (all ps > 0.11). The hippocampal correlation between the neural and the behavioural data was positive even when considering performance on the old (r = 0.27) and generalized (r = 0.54) objects separately, and it did not differ significantly across the two sets of trials (p = .34).

3.2.5. Missing evidence of increased representational similarity between objects belonging to the same category

Previous studies of categorization using a match-to-category task showed that in parietal (Freedman & Assad 2005; Jiang et al., 2007) and prefrontal (Jiang et al., 2007, 2018) cortices stimuli belonging to the same category were represented more similarly compared to those belonging to different categories. We asked whether the same effect could be observed in our experiment, even if we use a different task. We started investigating the left angular gyrus and the right hippocampus and observed that none represented within-category objects more similarly than across-categories ones (left AG: t(20) = −1, p = .32; right HPC: t(20) = −0.81, p = .43; t-tests)(Fig. 4D). We then looked at the other brain regions showing a common representational code between objects and their categorical names, observing the same pattern (left fusiform/vOT: t(20) = −0.67, p = .51; right SPL: t(20) = −2.48, p = .02; right IPL: t(20) = −4.86, p < .001; right AG: t(20) = −1.70, p = .10). Similarly, we looked at those brain regions representing the individual object identities, confirming the same trend (MFG: t(20) = −1.63, p = .12; IFG: t(20) = −2.03, p = .05; t-tests). A whole brain level searchlight confirmed (even when using a low threshold (p<0.05, FWE corrected at q<0.05)) no effect. None of these brain regions showed to represent categorical similarity between objects before learning (all p > .23), nor they showed to have significantly increased this representation post-learning compared to pre-learning (all p > .16).

4. Discussion

In this experiment we sought to investigate how the human brain categorizes multisensory stimuli using language. To this end, we acquired fMRI data before and after a behavioral training where participants learned that novel audiovisual objects belonged to 4 orthogonal
symbolically labelled categories. We found evidence of three different effects (potentially representing three representational stages of symbolic categorization): (i) increased sensitivity and specialization to the preferred sensory features in primary sensory regions, (ii) emergence of multisensory integration in the left angular gyrus, and (iii) emergence of abstract categorical representation common to objects and words in a series of associative cortices especially in the right hippocampus.

4.1. Sharpening and segregation of sensory features in sensory regions (Stage 1)

According to the classical models of non-symbolic categorization, following categorical training the sharpness of the neural response to categorized stimuli increases in those sensory regions where the specific sensory modality is processed, at least when unisensory stimuli are used (e.g., Kourtzi et al., 2005; Op de Beeck et al. 2006; Folstein et al. 2013; Brants et al., 2016; Jiang et al., 2007; Ley et al., 2014; Ley et al., 2012; Ahveninen et al., 2011; Bao 2015; Jiang et al., 2018). But what happens when more than one sensory modality is relevant for categorization? First, during our symbolic categorization task we showed a very similar sensory sharpening, where the two sensory features defining each object identity (size and sound) were separately processed in perceptual areas of the visual and acoustic pathways (LOC and aSTG - Rauschecker and Tian 2000; Obleser et al. 2007; Mishkin et al. 1983; Desimone and Schein 1987; Livingstone and Hubel 1988) and were more differentiated during the post-learning symbolic categorization task than during pre-learning. This effect is reminiscent of “differentiation” (e.g., Goldstone and Styers, 2001) and “acquired distinctiveness” (Lawrence 1949) effects, namely the ability to better discriminate two previously integrated dimensions (in our case size and pitch), or to better distinguish between the different levels in them (for instance in our case small size vs big size). Second, we observed that, concurrently, the same sensory regions decreased their sensitivity to differences along the non-preferred sensory modality: i.e., acoustic areas became less sensitive to differences in the size of the objects; visual areas became less sensitive to differences in the sound produced by the objects. We relate this finding to the behavioural evidence of ‘acquired equivalence’ effects, usually reported for dimensions that are task irrelevant (Waller 1970). In our specific task, the “irrelevance” of a sensory dimension should be intended as the fact that it is not the preferred (or matching) modality for a brain region. Similarly to our study, Lemus and colleagues (2010) previously recorded single neurons in macaque monkeys discriminating, on interleaved trials, between two tactile or two acoustic stimuli. While several neurons in the somatosensory cortices and primary auditory cortex responded to both visual and auditory stimuli, the stimulus identity could only be decoded from responses to their principal sensory modality. Thus, the authors suggested that during multisensory stimulation the representations of the different sensory modalities compete against each other, and sensory cortices select one over the other, according to their perceptual preference. The results of our experiment are congruent with this view, as we observed that within sensory regions information along the non-preferred sensory modality was reduced/suppressed (acquired equivalence), in favor of higher sensitivity for preferred sensory differences (acquired distinctiveness). These kinds of “suppressive” effects may be entirely overlooked in more classical multisensory stimulation experiments where the different sensory features of stimuli often do not orthogonally vary but, rather, are correlated and thus predictive of one another. In our experiment, on the contrary, the two sensory features varied orthogonally, such that allowing one modality to interfere with the encoding of the other would potentially reduce accuracy in stimulus recognition, the task that subjects were asked to perform. In this sense, the amount and type of multisensory integration that can be observed in sensory cortices might be crucially determined by the task and stimuli features used during the experiment at hand, and future work should further and directly investigate their specific role in influencing multisensory information coding in sensory areas. Clearly, the link of post-
to the processing of word meanings, showing that this region responds more strongly to real words compared to pseudowords (see Binder et al., 2009 for meta-analysis), and in particular its recruitment is stronger for high-frequency words (Graves et al., 2010) and for words that refer to concrete objects, compared to abstract ones (Binder et al. 2005). The central role of the left angular gyrus in processing the semantics of words has been summarised in a comprehensive meta-analysis of 120 neuroimaging studies (Binder et al., 2009), that reliably showed how, across different experimental paradigms and functional neuroimaging techniques (fMRI and PET), the left angular gyrus showed the most dense concentration of activation foci. Moreover, Fairhall & Caramazza (2013) showed similar multivariate activity in the left angular gyrus when people processed images of well-known real-life object categories (e.g. fruits, vegetables, tools, etc.) and their names. Consistently, in our study we showed the development of a common representational code between object categories and their names, thus between the concrete representation of a word meaning (the referent) and the word itself (the symbol). This result is in line with the proposal that at least part of the cortical territory devoted to the representation of object categories is recruited also to represent their symbolic form (Furhemmuller 2013): this has been proven true in the case of well known concepts or categories of objects, such as actions (Hauk et al., 2004), numbers (Piazza et al., 2007; Eger et al., 2009), colors (Simmons et al., 2007 Neuropsychologia), tools (Chao and Martin 1999) or places (Kumar et al., 2017). A key question for future studies is whether or not the effect in angular gyrus in our design is dependent on the specific sensory information used (audiovisual): a possible extension of our work would require the use of stimuli varying in different modalities, perhaps acoustic and tactile. A similar work by Levine & Schwarzbach (2017) has shown a single cortical locus where categorical identity of tactile-acoustic material could be recovered, in the right intraparietal sulcus, thus suggesting that the parietal cortex, broadly speaking, might play a crucial role in multisensory categorization. That study, however, did not investigate the influence of symbols and lexicalization during categorization.

4.4. The role of the right hippocampus

When we looked for signatures of the association between object categories and their names at the whole brain level, we revealed additional regions supporting this relation spanning both hemispheres: besides the left angular gyrus, which confirmed our previous results, the object-name correspondence was evident in the right hippocampus. Crucially, we observed strong categorical abstraction in the hippocampus, quantified in two different ways. First of all, we showed that a classifier trained to distinguish categories using hippocampal multivariate activity evoked by one stimulus format (e.g., objects, or words) could efficiently generalize to the other format (e.g., words, or objects). Second, we showed that hippocampal activity did not contain information about stimulus presentation modality, at least when this was quantified using a linear classifier. These results indicate a maximum level of abstraction, at least when compared to the left angular gyrus, that on the contrary did not show significant cross-decoding generalization, and was sensitive to presentation modality. While it would be possible to argue that the null finding of the presentation modality decoding in the hippocampus might be related to signal loss from the medial temporal lobe (Schmidt et al., 2005; Bellgowan et al., 2006; Olman et al., 2009), this hypothesis wouldn't fit with the positive results of the cross-modal correlation analysis and of the cross-format decoding approach. This leads to the more plausible explanation that the hippocampus might be fundamentally involved in (re)constructing the relevant association between object categories and their symbolic form providing a more abstract representational code. Previously, this region has been indicated as crucial in supporting the emergence of novel associations between object pairs or word pairs (Spiers et al., 2001; Giovanello & Keane 2003; Clark et al., 2018). Here we expanded those observations in showing that it is also involved in supporting novel word-object associations, similar to those observed during face-naming tasks (Sperling et al., 2003; Jackson & Schacter 2004). Although extreme caution should be adopted in drawing parallels with electrophysiological literature when working with functional MRI, the patterns of activity that we observed, with highly selective responses for different categories invariant to the stimulus presentation modality, reminds of the tuning properties of hippocampal “concept cells”, neurons that are highly selective to specific stimulus identities (e.g. pictures of a specific person or place) but highly invariant to their presentation format (Quiroga et al., 2005), which are believed to represent the neural scaffolding for new concepts to be rapidly acquired (Quiroga 2012). Interestingly, we observed that the degree of precision with which the hippocampus encoded the correspondence between object categories and the matching words was a significant predictor of the behavioral performance reached by subjects in a symbolic categorization task during the last day of training, when we presented participants not only with the objects they have been trained on, but also with novel exemplars, representing audiovisual combinations that they had never seen before. The role of the hippocampus in creating novel semantic memories, such as lexicalized categories of objects, had been overlooked. Our results suggest that this region is likely recruited to construct relations between any element of experience, thus moving beyond “simple” episodic associations of a specific, particular object exemplar with a particular word, but rather encoding the specific mapping rules that are used to generate the object-name associations. This is in line with recent clinical and theoretical works (e.g. Backus et al., 2016, Blumenthal et al., 2017, Mack et al., 2016, Zeithamova et al., 2012, Schlichting and Preston 2016, Bowman and Zeithamova 2018, Zeithamova et al., 2019), that challenges the classical view of a marginal hippocampal involvement in semantic memory. As predicted by early computational models (McClelland et al., 1995), the hippocampus, potentially in light of its high invariance to basic metric details (experimentally demonstrated, for instance, in Quiroga et al., 2005), might provide the optimal neural substrates for rapid associative coding of arbitrary contents of experience, that is then encoded in neocortical regions for later - and longstanding - recalling. This framework would predict, then, that the effect here observed might be transitory, at least with respect to two variables: the task (using a task that does not require the explicit association between object categories and their name should result in lower hippocampal involvement) and the time (the longer the experience with a particular category, the lower the need to recruit the hippocampus and the higher the involvement of long-term neocortical storage sites, such as the left angular gyrus). These predictions can be easily tested in follow-up experiments where task and training time are modulated accordingly.

The last aspect that remains surprising, for what concerns the role of the hippocampus, is its right lateralization. The right hippocampus is classically associated to visuo-spatial memory, which doesn’t bear any relevance to our experimental design, unless we interpret our novel semantic space as a “cognitive space” whose structure can be captured in a spatial format (Bellmund et al., 2018). This would not be a totally novel idea, and an increasing number of studies is showing how brain regions holding specific spatially-tuned representational code recruits the same mechanisms to represent more abstract non-spatial information (Constantinescu et al., 2016; Garvert et al., 2017; Theves et al., 2019; Viganò and Piazza 2020) in the form of an internal “cognitive map” (Tolman 1948; Behrens et al., 2018) of memories and concepts. Future studies should more directly address this possibility and explore whether and how these mechanisms are also crucial during semantic categorical learning.

4.5. The absence of similar representations for within-category objects

One unexpected result of our study was the absence of any categorical response when restricting to objects only. Typically, studies focusing on categorization employ non-verbal tasks that require the comparison of two objects, presented with a short delay between each other. Par-
ticipants are instructed to recall the categorical membership of the first stimulus before the second one is presented and then to indicate whether the second stimulus, presented after the short delay, belongs to the same category of the first one. Both electrophysiological (e.g. Freedman & Assad 2005) and fMRI (e.g. Jiang et al., 2007; 2018) studies using this procedure then analyze the activity evoked by the first stimulus of each pair and usually find that objects belonging to the same categories evoke similar activity when presented as first stimulus of a pair. In our task, this procedure was modified, to be closer to how humans usually access and express categorical memberships of real-world objects: through associating the objects with the symbolic label of their category (see Introduction). Thus, they were instructed to press a button when an object was preceded (or followed) by the corresponding name. We observed similar representations between objects and their names, as required by the task, but we did not observe the implicit similarity between two objects that are associated with the same name, as recovering that information was not required by the task. Why is this result relevant for the study of how animals, and humans in particular, categorize objects in the world? It is clear, from our experiment, that the brain mechanisms that support symbolic and non-symbolic categorization only partially overlap: while for the first stages of perceptual identification these coding schemes seem to be mostly consistent across tasks and domains (here we showed the predicted perceptual sharpening, and extended it to the domain of multisensory categorization), for the later stages of categori-

cal decision-making there seems to be a more substantial variability that depends on the task. Indeed, as stated in Jiang et al., 2018 referring also to previous works (Jiang et al., 2007; Roy et al., 2010; Van der Linden et al. 2014), the categorical effects reported during non-symbolic catego-

rization are strongly modulated by what participants are instructed to do: if they are actively engaged in recognizing categorical similarities between objects (such as during a “delayed match to category task”), then the within-category similarities emerge, but if they are asked to process the same stimuli without paying attention to categorical mem-

berships (e.g. during spatial localization, Jiang et al., 2007), these effects disappear. In our task, indeed, participants were not asked to compare object exemplars, but to recover, for each one of them, the correct cat-

ergy name. As discussed in Lupyan 2012, referring also to the work of McMurray & Spivey 2000, and McMurray et al. 2008, “the reason that within-category differences are never fully collapsed is that doing so would render the representations useful only for that single type of categorization. This is never the case. So, e.g., we need to know not only whether something is a car, but whether it is our car, whether it is moving, and whether it poses a present danger”. Comparing our re-

sults to those of previous studies using different tasks, then, it seems that the neural representation of objects during categorization is highly affected by what is relevant for the task at hand, demonstrating how the brain can flexibly adapt to meet what is required by the interaction with the environment. Our null result can thus be seen as the consequence of the hybrid task adopted, allowing assessment of categorical knowledge, yet not requiring object-to-object comparison. This setting might weaken the effects of category similarity, and, together with intrinsic limitations in the analytical approach employed, hamper our sensitivity to those effects.

4.6. Limitations and future perspective

Far from being conclusive, our study suffers from some limitations that impose caution in the interpretation and generalization of the results.

First of all, although offering rigorous control of the intervening vari-

ables, our stimulus space and object categories are highly artificial. Fu-

ture studies aiming at generalizing our findings to real world scenarios, may wish to adopt more ecological stimuli, such that of real objects and known words. This however would conflict with the fact that par-

ticipants will have already acquired and established categorical knowl-

dge of real-life objects. One possibility to avoid this would be that of adapting the design to children and tracing their brain changes while they learn: although methodologically more challenging given the age of participants, this approach would be highly informative on how natu-

ral categories are acquired by means of language and how words acquire meaning. Alternatively, adult participants can be trained and tested us-

ing material coming from fictional stories, such as creatures coming from science-fiction novels (e.g. alien species from the Star Wars uni-

verse). Clearly, in these scenarios it would be impossible to control the variation of perceptual features across stimuli and categories in a para-

metric way as we did in our experiment, thus potentially omitting more subtle effects that we were able to unveil, such that of multisensory seg-

regation.

The second partial limitation of the current study regards the task we adopted. Studying symbolic categorization, namely the ability to categorize an object by means of its label or name, is intrinsically an association task. The role of the hippocampus in associative and rela-

tional memory is highly recognized (e.g. Eichenbaum and Cohen 2001), therefore our results might be seen as just another example of its role in associative memory. However, three key findings here discussed go beyond a simple associative framework: i) the hippocampus was not the only region showing this associative property; ii) compared to the other regions, the hippocampus showed higher abstraction in representing this association; and iii) the precision by which object categories were repre-

sented similarly to their category names in hippocampus was predictive of categorization performance on the test day, when also novel objects were presented to subjects, thus indicating that this region is perhaps representing something more profound that a simple episodic associa-

tion. Our study thus is in line with a growing body of evidence indic-

ating that the hippocampal formation is involved, above and beyond associative memory formation, in the acquisition and representation of conceptual knowledge (Blumenthal et al., 2017; Mack et al., 2016; Zeithamova et al., 2012; Schlichting and Preston 2016; Bowman and Zeithamova 2018; Zeithamova et al., 2019; Bellmund et al., 2018).

Finally, with our current design we could not properly investigate the brain correlates of learning to categorize objects using symbols, that is, we did not monitor brain activity during training. Although challeng-

ing and time consuming, this is a promising idea for follow-up experi-

ments, and will potentially elucidate the putative transitory role of the hippocampus during the early stages of learning.

4.7. Symbolic categorization: concluding remarks

It takes few instants to recognize an object in our surroundings and to select the correct behavioural scheme to interact with it (a lion: rush away, a friend: approach smiling). Humans use language to recognize objects and their category and to communicate about them. In this study we showed that symbolic categorization comprises at least three repre-

sentational stages, that are only partially distinguishable at the brain level. When we are asked to categorize an object by means of its lin-

guistic label, (i) our sensory regions selectively process the sensory fea-

tures that are relevant for categorization, (ii) the left angular gyrus in-


tegrates the different sensory features into unique object identities and connect them to the correct name, and (iii) the hippocampus supports this relation by encoding the abstract associative rule. Future studies should investigate the temporal course of these different stages, and di-

rectly compare the effects of categorizing objects using words vs doing it without language, to describe the major advantages of symbolic cate-

gorization in acquiring knowledge about the world.

Data and code availability statement

Following the guidelines of the University of Trento on data shar-

ing and privacy, data and code used for the present study are available from the corresponding author only for purposes related to the original research question.
Acknowledgments

SV and MP designed the experiment; SV collected the data; SV and VB analysed the data; SV, VB and MP wrote the manuscript. The authors thank Maria Varela for her help in data collection.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1101/2018.08.07.17.81939.
