

# Accepted Manuscript

Neural pattern similarity underlies the mnemonic advantages for living words

Xiaoqian Xiao, Qi Dong, Chuansheng Chen, Gui Xue



PII: S0010-9452(16)30054-5

DOI: [10.1016/j.cortex.2016.03.016](https://doi.org/10.1016/j.cortex.2016.03.016)

Reference: CORTEX 1711

To appear in: *Cortex*

Received Date: 15 July 2015

Revised Date: 8 January 2016

Accepted Date: 16 March 2016

Please cite this article as: Xiao X, Dong Q, Chen C, Xue G, Neural pattern similarity underlies the mnemonic advantages for living words, *CORTEX* (2016), doi: 10.1016/j.cortex.2016.03.016.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Neural pattern similarity underlies the mnemonic advantages for living words**

Xiaoqian Xiao<sup>1,2</sup>, Qi Dong<sup>1,2</sup>, Chuansheng Chen<sup>3</sup>, Gui Xue<sup>1,2\*</sup>

<sup>1</sup> State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, PR China

<sup>2</sup> Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, 100875, China

<sup>3</sup> Department of Psychology and Social Behavior, University of California, Irvine, 92697, United States

**Abbreviated title: Mnemonic advantages for living words**

**Contribution**

G.X. designed the experiment. X. X. performed the study. X. X. and G.X. analyzed the data. X.X., Q.D., C.C. and G.X. wrote the manuscript.

**Competing financial interests**

The authors declare no competing financial interests.

**Corresponding author:**

**Gui Xue: [guixue@gmail.com](mailto:guixue@gmail.com)**

**Abstract**

It has been consistently shown that words representing living things are better remembered than words representing nonliving things, yet the underlying cognitive and neural mechanisms have not been clearly elucidated. The present study used both univariate and multivariate pattern analyses to examine the hypotheses that living words are better remembered because (1) they draw more attention and/or (2) they share more overlapping semantic features. Subjects were asked to study a list of living and nonliving words during a semantic judgment task. An unexpected recognition test was administered 30 minutes later. We found that subjects recognized significantly more living words than nonliving words. Results supported the overlapping semantic feature hypothesis by showing that (a) semantic ratings showed greater semantic similarity for living words than for nonliving words, (b) there was also significantly greater neural global pattern similarity (nGPS) for living words than for nonliving words in the posterior portion of left parahippocampus (LpPHG), (c) the nGPS in the LpPHG reflected the rated semantic similarity, and also mediated the memory differences between two semantic categories, and (d) greater univariate activation was found for living words than for nonliving words in the left hippocampus (LHIP), which mediated the better memory performance for living words and might reflect greater semantic context binding. In contrast, although living words were processed faster and elicited a stronger activity in the dorsal attention network, these differences did not mediate the animacy effect in memory. Taken together, our results provide strong support to the overlapping semantic features hypothesis, and emphasize the important role of semantic organization in episodic memory encoding.

**Keywords**

Animacy effect; Neural global pattern similarity; MTL; Global match; Representational similarity analysis; Episodic memory

## Introduction

Recent studies have shown that living things are better remembered than nonliving things (Bonin, Gelin, & Bugaiska, 2014; Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013; VanArsdall, Nairne, Pandeirada, & Cogdill, 2014; VanArsdall, Nairne, Pandeirada, & Blunt, 2013). This animacy effect of memory has been consistently demonstrated in both recognition (VanArsdall et al., 2013) and free recall (VanArsdall et al., 2013, 2014), and when using either words (Nairne et al., 2013) or pictures (Bonin et al., 2014) as the learning materials. Besides their mnemonic advantages, living things are also easier to recognize (Lag, 2005; Turnbull & Laws, 2000), name (Laws, Leeson, & Gale, 2002; Laws & Neve, 1999), and categorize (Turnbull & Laws, 2000) than are nonliving things. Nevertheless, the cognitive and neural mechanisms underlying the animacy effect of memory are poorly understood.

There exist two main possible explanations of the animacy effect. One possibility is that living stimuli are better remembered because they draw more attention than nonliving stimuli (Lipp, Derakshan, Waters, & Logies, 2004; New, Cosmides, & Tooby, 2007) and attention is important for the formation of lasting memory (Chun & Turk-Browne, 2007; Kim, 2011; Turk-Browne, Golomb, & Chun, 2013; Uncapher & Wagner, 2009). This attentional bias for living stimuli can be further traced to their evolutionary advantage for human survival and fitness (Bonin et al., 2014; Laiacona, Barbarotto, & Capitani, 2006; Laws, 2000; VanArsdall et al., 2014), which has also been invoked to explain a documented gender-by-category interaction (males are better in naming and recognizing nonliving things whereas females are better at processing living things) (Laws, 1999; Laws 2004). Further supporting the deep biological root of the sensitivity to living things, research has shown that children develop their knowledge of living objects at a very early age (Lambon Ralph, Graham, Ellis, & Hodges, 1998; Rakison & Poulin-Dubois, 2001; Rostad, Yott, & Poulin-Dubois, 2012; Silveri, Cappa, Mariotti, & Puopolo, 2002) and their preference for living things seems to be innate because brain damage due to posterior cerebral artery infarctions at one day of age leads to specific impairment in the processing of living things (Farah & Rabinowitz, 2003).

An alternative perspective is the overlapping semantic features hypothesis, which emphasizes the greater overlap in semantic features among living objects than among nonliving objects (Cree & McRae, 2003; Dixon, Bub, & Arguin, 1998; Garrard et al., 2001; Humphreys & Forde, 2001; McRae, Cree, Seidenberg, & McNorgan, 2005; McRae, de Sa, & Seidenberg, 1997; Sartori, Gnoato, Mariani, Prioni, & Lombardi, 2007; Sartori & Lombardi, 2004; Sartori, Lombardi, & Mattiuzzi, 2005; Zannino, Perri, Pasqualetti, Caltagirone, & Carlesimo, 2006a, 2006b; Zannino, Perri, Pasqualetti, Di Paola, et al., 2006). Due to the tight connection between semantic memory and episodic memory, the overlapping semantic features could affect episodic memory via a global matching mechanism (Criss & Shiffrin, 2004; McClelland, McNaughton, & O'Reilly, 1995; Norman, 2010; O'Reilly, Bhattacharyya, Howard, & Ketz, 2014; Shiffrin, Huber, & Marinelli, 1995). In particular, it is proposed that items with greater global similarity (or shorter semantic distance) with all other studied items have stronger episodic memory strength (Clark & Gronlund, 1996).

Consistent with the global matching hypothesis, researchers have recently used representational similarity analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008) to study the relationship between neural global pattern similarity and memory. The basic idea of RSA is to examine the geometric relations (e.g., distance) of materials in certain representational space as well as across different levels of analyses (e.g., at the behavioral, computational, or neural level) (Khaligh-Razavi & Kriegeskorte, 2014; Kriegeskorte, Mur, Ruff, et al., 2008; Su, Zulfiqar, Jamshed, Fonteneau, & Marslen-Wilson, 2014). For example, previous studies have found that items with high neural pattern similarity as measured by fMRI (Davis, Xue, Love, Preston, & Poldrack, 2014; LaRocque et al., 2013) or EEG (Lu, Wang, Chen, & Xue, 2015) were associated with better recognition memory. Specifically, Davis et al. (2014) found that neural global pattern similarity in the MTL, which reflects the representational overlap with other studied items, was correlated significantly with recognition confidence of words. Using high-resolution fMRI that enables the discovery of finer functional dissociations in the MTL, LaRocque et al. (2013) found that better memory was associated with greater

neural global pattern similarity in both perirhinal cortex and parahippocampal cortex, but with greater pattern distinctiveness in the hippocampus.

The current study used fMRI to test the attentional bias hypothesis and the overlapping semantic features hypothesis. Based on the attentional bias hypothesis, we would expect that (a) compared to nonliving words, living words would be processed with greater attention as shown by greater activation in attention-related brain regions, such as the dorsal frontoparietal network (for a review, see Corbetta et al., 2008; also see Kim, 2011; Uncapher & Wagner, 2009) and (b) the differences in brain activation between living vs. nonliving words would explain (i.e., mediate) the animacy effect of memory. In contrast, based on the overlapping semantic features hypothesis, we would expect that compared to nonliving words, living words would show a greater overlap in semantic features and in neural representations (i.e., greater neural global pattern similarity), which would in turn explain (i.e., mediate) the animacy effect in memory. Because we wanted to account for the differences in episodic memory of living vs. nonliving words, we focused on brain regions where the activation level or global neural pattern similarity could predict subsequent memory.

## **Materials and Methods**

### *Participants and experiment design*

Detailed information regarding the participants, experimental design, behavioral data analysis, and functional magnetic resonance imaging (MRI) data preprocessing can be found in Supplementary Material of our previously published paper (Xue et al., 2010). Briefly, 22 subjects (11 males, mean age =  $19.56 \pm 1.76$  years, ranging from 17 to 25 years) participated in this study. They studied 60 high-frequency Chinese nouns (with half representing living objects (e.g. bee, ant, frog and monkey) and half representing nonliving objects (e.g. chair, gate, match and pillar)) in the scanner over three scanning runs. The living words and nonliving words were matched in terms of word frequency (living:  $4560 \pm 15.20$ /million; nonliving:  $4545 \pm 21.24$ /million,  $t(58) = 0.014$ ,  $p = .989$ ) and stroke number (living:  $15.17 \pm 5.24$ ; nonliving:  $15.73 \pm 4.21$ ,  $t(58) = -0.462$ ,  $p = .646$ ).

Each item was repeated three times. A slow event-related design (12 s for each trial) was used in this study (Fig. 1A) to obtain better estimates of single-trial BOLD response. The trial started with a 1 s fixation point, followed by a Chinese word that was presented on the screen for 3 s (or until a response was made, whichever came first). Subjects were asked to make a semantic judgment (i.e. living or nonliving) about the word by pressing a button with their left or right thumb. To prevent further encoding of the word, three seconds after the onset of the word, subjects were asked to perform a perceptual judgment task (to judge the orientation of a Gabor image tilting 45° to the left or the right) for 8 s. A self-paced procedure was used to make this task engaging.

Thirty minutes after the scan, subjects were asked to perform two unexpected memory tests. In the first test (i.e. the free recall test), subjects were asked to write down the words they had studied in the scanner, regardless of the order of presentation. After the free recall test, they were given a recognition test and asked to judge whether a word was old or new on a 6-point scale, with 1 indicating definitely new and 6 indicating definitely old. In this test, 60 original words and 60 foils (30 living words and 30 nonliving words) were randomly mixed and presented one by one. The present analysis focused on the recognition test because the global match hypothesis was mainly developed to account for recognition memory performance (Norman, 2010).

#### fMRI image data acquisition

Scanning was carried out in the MRI Center of Beijing Normal University using a 3.0T Siemens MRI scanner with an 8-channel head coil. Functional data were acquired using T2\*-weighted EPI (FOV = 200 \* 200mm; matrix = 64 \* 64; slice thickness = 4mm; TR/TE/θ = 2s/30ms/90°). Thirty contiguous axial slices parallel to the AC-PC line were obtained to cover the whole cerebrum and partial cerebellum. High-resolution structural images using gradient-echo pulse-sequence were acquired for the whole brain (FOV = 256 \* 256mm; matrix = 192 \* 256; slice thickness = 1.33mm; TR/TE/θ = 2530ms/3.39ms/7°).

#### Behavioral data analysis

Reaction time and accuracy data were collected and analyzed according to the subsequent memory performance and the stimuli's semantic categories. For memory performance, items recognized with high confidence (scored 6) were defined as remembered items (R) and those scored 4 and lower were defined as forgotten items (F). Both the ratio of recognized items and  $d'$  were used to measure memory performance. Paired  $t$  test was conducted to examine the effect of semantic category on subsequent memory performance. Repeated measure ANOVA was used to examine behavioral data from the encoding task.

#### Semantic similarity rating

Eight additional native Chinese college students (four males) were recruited to rate the pair-wise semantic similarity of the 60 words on a 7-point scale. Subjects were told to rate the similarity of each pair of words based on their meanings. There were 1770 pairs in total, which were randomly separated into 9 groups. Subjects took a short break after finishing a group. Second-order similarity (Islam & Inkpen, 2006) was calculated by computing correlations between the raw rating vectors of two items. The rationale behind the use of second-order similarity is that words that are similar to each other should show similar profiles of relations with all other items. This second-order similarity measure has been shown to reduce within-subject noise in the ratings, and to be more reliable than the first-order similarity index (Davis et al., 2014).

#### fMRI imaging preprocessing

Image preprocessing and statistical analyses were carried out using FEAT (fMRI Expert Analysis Tool) version 5.98. The first three volumes before the task were automatically discarded by the scanner to allow for T1 equilibrium. The remaining images were then realigned, spatially smoothed (only for activation-based analysis) with a 5mm full-width-half-maximum (FWHM) Gaussian kernel, and temporally filtered (nonlinear highpass filter with a 90s cut-off). The EPI images were first registered to the magnetization-prepared rapid gradient echo (MPRAGE) structural image and then into the standard Montreal Neurological Institute (MNI) space, using affine transformations. Registration from structural images to the standard space was further refined using



FNIRT nonlinear registration.

### Activation-based analysis

As mentioned in Introduction, this study focused on the animacy effect in memory, so we first identified brain regions whose activation was related to episodic memory. We examined the subsequent memory effect (SME) using the GLM model implemented within the FILM module of FSL. The recognized and forgotten words were separately modeled. Trials scored 5 from the recognition task were treated as a nuisance variable. The incorrect trials from the perceptual orientation task were coded as an additional nuisance variable, whereas the correct orientation trials were not coded and thus were treated as an implicit baseline. The SME was defined as the differences between recognized and forgotten words. A higher-level analysis involved cross-run contrasts for each subject using a fixed-effects model. These contrasts were then used for group analysis with a random-effects model, using FLAME (FMRIB's Local Analysis of Mixed Effect) stage 1 only with automatic outlier detection (Beckmann, Jenkinson, & Smith, 2003; Woolrich, 2008; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004). Because we were interested in the activation differences between living and nonliving words in the brain regions that showed the SME, we used a relatively liberal threshold to find these regions ( $z > 2$  and a cluster probability of  $p < 0.05$ , corrected for whole-brain multiple comparisons using Gaussian random field theory).

Then we tested whether living words showed greater neural activation than nonliving words within the SME-sensitive ROIs. Because of our special focus on the attention network, we also defined two masks in the dorsal attention network, including the dorsal frontal cortex and dorsal parietal cortex, based on the Harvard-Oxford probabilistic atlas (25% threshold, 2mm). The dorsal frontal cortex included bilateral precentral sulcus and the middle frontal gyrus, while the dorsal parietal cortex included both the intraparietal sulcus and superior parietal lobule (Corbetta, Patel, & Shulman, 2008; Kim, 2011; Uncapher & Wagner, 2009). Signals within the SME-sensitive ROIs and the dorsal frontal and parietal ROIs were extracted from single-item estimation, separately for each subject and each item. Mixed-effects model (see below) was then performed with

activation level as the dependent variable and animacy category (living vs. nonliving) as the predictor, and subjects as a random effect. Finally, mediation effect was tested to see whether higher activation in living words within SME-sensitive ROIs could mediate the animacy effect in memory.

#### Mixed-effects model

Mixed-effects modeling is a powerful statistical tool that offers many advantages over conventional t test, regression, and ANOVA in sophisticated fMRI designs (Mumford & Poldrack, 2007; Ward, Chun, & Kuhl, 2013), especially when the number of trials differs by condition and/or across participants (e.g., in this study, participants remembered different numbers of words). In this study, the mixed-effects model was implemented with lme4 in R (Bates D, 2012). We used the likelihood ratio test to compare the models (with vs. without the predictor) to determine the effect of the predictor.

#### Mediation analysis

We performed the mediation effect test to examine whether the neural indices (activation level or nGPS) mediated the animacy effect in memory performance. We examined the relationship between 1) semantic category and memory performance ( $Y = a_1 + b_1X + \varepsilon_1$ ); 2) semantic category and neural response (activation level or nGPS) ( $M = a_2 + b_2X + \varepsilon_2$ ); and 3) semantic category and memory performance with mediator ( $Y = a_3 + b_3X + bM + \varepsilon_3$ ). In the equations, Y is the dependent variable (memory performance: remembered or forgotten; binary variable), X is the predictor (semantic category: living or nonliving; binary variable), and M is the mediator (neural indices, activation level or nGPS in separate models; continuous variable). The indirect effect was estimated as  $b_2 * b$ . We used distribution-of-the-product method to compute CIs; this method has been proved to be more accurate when the sample size is small (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Mediation effect size ( $R^2_{ind}$ ) was also presented according to the equation  $R^2_{ind} = R^2_{tot} - R^2_{dir}$  (Fairchild, Mackinnon, Taborga, & Taylor, 2009).

#### Neural global pattern similarity analysis

For pattern similarity analysis, we first estimated the single-item response. GLM was performed to compute the  $\beta$ -map for each of the 60 unique word stimuli. All three presentations of each stimulus were used as one single regressor, in which the presentation of each stimulus was modeled as an impulse and convolved with a canonical hemodynamic response function (double gamma) (Mumford, Turner, Ashby, & Poldrack, 2012).

Again we first identified the SME-sensitive brain regions using the searchlight method (Kriegeskorte, Goebel, & Bandettini, 2006) as described in Davis et al. (2014). For each voxel, signals were extracted from the cubic ROI containing 125 surrounding voxels. For each item, averaged pair-wise correlation was calculated between the item and all other words. These similarity scores were transformed into Fisher's z scores. Instead of using a parametric analysis to examine the relationship between memory strength and neural global pattern similarity (nGPS) (Davis et al., 2014), here we compared the differences between recognized (scored 6) and forgotten words (scored 1 to 4) to be consistent with the activation-based analysis.

We focused on the MTL, as its neural activation pattern has been suggested to reflect global match (LaRocque et al., 2013; O'Reilly et al., 2014). We also anatomically defined the anterior temporal lobe (ATL) and the middle temporal gyrus (MTG) (Binney, Embleton, Jefferies, Parker, & Ralph, 2010) because their involvement in semantic representation. The MTL (bilateral hippocampus and the parahippocampus gyrus), ATL (bilateral temporal pole and the anterior part of the superior temporal gyrus), and MTG (bilateral middle temporal gyrus) were defined based on the Harvard-Oxford probabilistic atlas included in the FSL package (threshold at 25% probability), and were used for small-volume correction.

Focusing on the regions that showed the SME in nGPS, we further examined whether nGPS differed between living and nonliving words. Similar to the activation-based analysis, the  $\beta$  value was extracted from each voxel within each ROI, separately for each word and each subject. The nGPS within each ROI was calculated using the

methods described above. Mixed-effects model was then applied with nGPS as the dependent variable to examine the animacy effect on nGPS. Finally, mediation effect test was performed to test whether higher nGPS in living words within SME-sensitive ROIs could mediate the animacy effect in memory.

## Results

### Behavioral results

#### Memory performance of living and nonliving words

As predicted, subjects recognized with high confidence significantly more living words (mean = 79.24%, SD = 13.01) than nonliving words (mean = 62.58%, SD = 16.81;  $t(21) = 6.640, p < .001$ ). The overall false alarm rate was low, 5.99% (SD = 4.79). However, consistent with the global match hypothesis, the false alarm rate was significantly higher for living than nonliving words ( $t(21) = 2.625, p = .016$ ). The unbiased index  $d'$  for living words (mean = 2.41, SD = 0.47) was also significantly higher than that for nonliving words (mean = 2.11, SD = 0.46;  $t(21) = 2.646, p = .015$ ) (Figure 2A).

#### Behavioral performance during encoding

We performed semantic category (nonliving vs. living) by repetition (first rep vs. second rep vs. third rep) by subsequent memory (remembered vs. forgotten) three-way repeated measure ANOVA on reaction time and accuracy during encoding. Reaction time was shorter for living words than for nonliving words ( $F(1,21) = 8.723, p = .008$ ). There was a significant repetition priming effect ( $F(1,21) = 46.492, p < .0001$ ). However, there was no significant subsequent memory effect for RT ( $F(1,21) = 1.435, p = .244$ ). Also, there was not any significant interaction effect (all  $p$ 's  $> .3$ ) (Figure 2B).

Accuracy was high and did not differ across conditions (mean = 0.976, SD = 0.034). Three-way repeated measure ANOVA revealed no significant effect of subsequent memory performance ( $F(1,21) = 2.345, p = .141$ ), semantic category ( $F(1,21) = 0.057, p = .814$ ), or repetition priming ( $F(1,21) = 0.728, p = .489$ ). There was also no significant interaction effect (all  $p$ 's  $> .2$ ) (Figure 2C).

### Semantic distance differences between living and nonliving words

As expected, compared to the raw ratings of (first-order) similarity, the second-order similarity was more reliable: The intra-class correlations (cross-subjects) were 0.65 and 0.77, for first- and second-order similarity, respectively. Using the second-order similarity (Figure 3A), the three-dimensional scaling analysis of pair-wise similarities captured a clear semantic category structure (Figure 3B). We then calculated the global semantic similarity of each word by averaging the z-transformed pair-wise similarity between a given word and all other words. This analysis revealed significantly higher global semantic similarity for living words than for nonliving words ( $t(29) = 4.097$ ,  $p = .0003$ ; Figure 3C). Further analysis indicated that the higher global semantic similarity for living words was caused by the significantly higher within-category similarity ( $t(29) = 7.236$ ,  $p < .0001$ ; Figure 3C), whereas the between-category similarity was comparable ( $t(29) = 0.0008$ ,  $p = .9994$ ; Figure 3C).

### **fMRI results**

#### The attention network and the animacy effect in memory

The attentional bias hypothesis predicted that activations in the attention network would show the animacy effect, which would in turn explain/mediate the animacy effect in memory. In order to test these hypotheses, we first identified brain regions with the subsequent memory effect (SME) by contrasting remembered vs. forgotten words. Whole-brain analysis revealed several brain regions showing the SME, including the left orbital frontal cortex (LOFC,  $-20, 14, -22$ ,  $Z = 3.17$ ), left hippocampus (LHIP,  $-30, -16, -22$ ,  $Z = 3.16$ ), the anterior segment of left parahippocampus gyrus (LaPHG,  $-16, 0, -24$ ,  $Z = 2.86$ ), left occipital cortex (LLOC,  $-24, -66, 18$ ,  $Z = 3.43$ ), and right temporal occipital fusiform (RFG,  $48, -58, -18$ ,  $Z = 3.27$ ) (Figure 4). However, no significant effect was found in the dorsal attention network.

Focusing on regions showing the SME, we then examined the animacy effect with a mixed-effects model. In this model, activation level was the dependent variable and semantic category was the predictor, and subjects were included as random effects. The results revealed that living words showed significantly higher activation in LHIP

( $\chi^2(1) = 8.418, p = .0185$ , after Bonferroni correction for the number of regions). The LaPHG ( $\chi^2(1) = 5.496, p = .095$ , corrected) showed a trend of significant differences after correction. No significant difference was found in RFG, LFOC and LLOC (all  $p$ 's > .15, uncorrected) (Fig. 4A).

Finally, a mediation analysis (Tofighi & MacKinnon, 2011) was conducted to examine whether the greater activation level of living words mediated the animacy effect in memory performance. This analysis revealed a significant mediation effect in LHIP (product of coefficients = 0.045, SE = 0.022, 95% CI is [0.008, 0.092], effect size = 0.041) (Fig.4B), but not in the LaPHG (product of coefficients = 0.034, SE = 0.022, 95% CI is [-0.004, 0.082]).

Although the above whole-brain analyses did not support the attentional bias hypothesis, ROI analyses focusing on the dorsal attention network would provide a more stringent test. Additional small volume correction was performed in the dorsal frontal and parietal cortices (see materials and methods). Results showed a higher activation level for living words than for nonliving words in the left dorsolateral prefrontal cortex (DLPFC, -52, 14, 32,  $Z = 3.53$ ), but the mixed-effects model showed no significant SME ( $\chi^2(1) = 2.225, p = .136$ ) and no significant mediation ( $\chi^2(3) = 1.837, p = .607$ , uncorrected).

#### Neural GPS in LpPHG and the animacy effect in memory

The overlapping semantic features hypothesis predicted that nGPS would show the animacy effect, which in turn would explain the animacy effect in memory. In addition, nGPS was expected to be correlated with semantic distance. Similar to the above analyses on activations, we focused on brain regions showing the SME. Using a searchlight procedure with a focus on the MTL, ATL and MTG (Fig. 5A), we found significant SME (remembered > forgotten) in nGPS in the left hippocampus (LHIP, -20, -30, -8,  $Z = 3.79$ ) and the posterior portion of the left parahippocampus (LpPHG, -32, -34, -16,  $Z = 3.68$ ) (Fig. 5B). These results were similar to those found using parametric analysis (Davis et al., 2014). No significant clusters were found in the ATL or MTG.

These results were consistent with previous studies that found no semantic representation in these regions (Bedny & Thompson-Schill, 2006; Binney et al., 2010; Copland et al., 2003; Devlin, Matthews, & Rushworth, 2003; C. L. Liu et al., 2006). It should be noted that the susceptibility artifacts (i.e., artifacts due to differences in magnetic susceptibilities of materials, especially around air-tissue interfaces) might have reduced the signal-to-noise ratio in these regions, thus reducing the power to detect the effect.

The mixed-effects model revealed a strong animacy effect of nGPS (living > nonliving) in the LpPHG ( $\chi^2(1) = 6.874, p = .009$ ), but not in the LHIP ( $\chi^2(1) = 2.106, p = .147$ , uncorrected) (Fig. 5C). Moreover, the above findings did not change even when the activation level in the respective region was added as a covariate (LpPHG:  $\chi^2(1) = 7.248, p = 0.007$ ; LHIP ( $\chi^2(1) = 2.067, p = 0.149$ , uncorrected). Further analysis revealed that the higher nGPS in LpPHG for living words was primarily driven by their higher within-category similarity ( $\chi^2(1) = 12.204, p < .001$ ), whereas living and nonliving words showed comparable between-category similarity ( $\chi^2(1) = .305, p = .581$ ) (Fig. 5C).

We further examined the association between nGPS and semantic global similarity and found that nGPS in the LpPHG was strongly and positively associated with semantic similarity ( $\chi^2(1) = 76.033, p < .0001$ ), suggesting that the nGPS reflects the semantic similarity. Finally, mediation analysis revealed that nGPS in the LpPHG significantly mediated the animacy effect in memory performance (product of coefficients = 0.028, SE = 0.016, 95% CI is [0.003, 0.065], effect size = 0.038) (Fig. 5D).

## Discussion

The current study aimed at examining the neural mechanisms underlying the animacy effect in memory, i.e., living words are better remembered than nonliving words. In particular, we tested whether this effect could be attributed to higher attention and/or greater neural global pattern similarity (nGPS) for living words than nonliving words because both attention and nGPS have been linked to better memory performance (See Introduction). Our results provide strong support for the overlapping representation

account but weak support for the attention account, at least under the current experimental condition.

Existing studies have consistently revealed greater semantic similarity for living words than for nonliving words (Cree & McRae, 2003; Dixon et al., 1998; Garrard et al., 2001; Humphreys & Forde, 2001; McRae et al., 2005; McRae et al., 1997; Sartori et al., 2007; Sartori & Lombardi, 2004; Sartori et al., 2005; Zannino, Perri, et al., 2006a, 2006b). To generate the semantic similarity measures, researchers have asked subjects to produce the features they think are central to a given concept (McRae et al., 2005). These feature norms (lists of features and corresponding frequencies) were then subjected to different algorithms (e.g., cosine correlation or  $\chi^2$  distance) to calculate semantic similarity across words. Because such feature norms are not available for Chinese words, the present study measured semantic similarity based on participants' subjective ratings of semantic similarity for each pair of concepts. Multi-dimensional scaling revealed a clear structure for animacy categories, with living words clustered more closely with one another than nonliving words. Although we could have used the feature norms from the English translations of the Chinese words, there does not exist an exact word-by-word correspondence between the two languages. Furthermore, the feature norm approach in general does not consider the weight of each feature's contribution to the overall similarity, nor does it consider the interaction between features, which might undermine the calculation of semantic similarity.

Although living things are overall more similar to each other than are nonliving things, researchers have also found opposite results in some special cases. For example, nonliving things showed greater within-category (structural) similarity than did living things when pictures were used (Gale & Laws, 2006; Laws & Gale, 2002; Laws & Neve, 1999; Marques, Raposo, & Almeida, 2013) or when particular subcategories of living (i.e., body parts) and nonliving things (i.e., musical instruments) were used (Laws, Gale, Frank, & Davey, 2002). Nevertheless, these studies converged with the other studies to show that higher similarity (either semantic or structural similarity) is associated with more processing errors (i.e. naming) but better memory (Laws & Gale, 2002; Zannino,



Perri, Pasqualetti, Di Paola, et al., 2006). These results provide further evidence to support the global matching hypothesis, which would predict that the overlap of representations during encoding would lead to better memory. The current study focused on semantic features because a semantic judgment task using words as stimuli was used. Future research should use different encoding tasks, and different types of stimuli (words vs. pictures) from certain semantic subcategories, to further test the global matching hypothesis. These studies would also help to specify the representation features that contribute to the global similarity measures and memory performance.

Furthermore, we found that the subjective ratings of similarity were associated with nGPS. Specifically, we used representational similarity analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008) and found that nGPS was correlated with subjective similarity across word pairs and that nGPS was greater for living words than for nonliving words. This finding is consistent with the growing evidence that neural activity patterns could reflect mental representations (Kriegeskorte & Kievit, 2013). For example, based on multivoxel pattern analysis, brain activity patterns have been found to reflect behavioral category judgment of real-world materials (Hiramatsu, Goda, & Komatsu, 2011), category structure of animate concepts (Connolly et al., 2012), behavioral judgment of perceptual similarity (Mur et al., 2013), as well as the individuals' unique perceptual judgment (Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014). Our results add to this literature by showing that neural activity patterns may reflect the semantic representations of living and nonliving words.

More importantly, our results clearly suggest that nGPS provides a mechanistic understanding of the mnemonic advantages of living concepts. Consistent with the global match theory (for a review, see Clark and Gronlund, 1996), the current study corroborates recent neuroimaging evidence that a higher level of overlapping representations benefits memory performance (Davis et al., 2014; LaRocque et al., 2013; Lu et al., 2015). We found that living words elicited greater nGPS in LpPHG than did nonliving words, and that the nGPS in this region significantly mediated the animacy effect in memory. This mediation effect was robust even after we added the activation

level in this region and LHIP activity (which also mediated the animacy effect, see below) as covariate factors.

The parahippocampal gyrus (PHG) has been consistently implicated in recognition memory. In particular, recent representational-hierarchical account argues that the highly conjunctive representation in PHG is required to differentiate confusable items (Barens et al., 2012; Bussey & Saksida, 2005; Cowell, Bussey, & Saksida, 2006; McTighe, Cowell, Winters, Bussey, & Saksida, 2010). This representational-hierarchical view states that MTL cortex represents highly complex conjunctive features. When MTL cortex is intact, conjunctive representations make a given memory distinctive from interferences. When MTL cortex is impaired, judgments have to be made based on highly susceptible simple feature representations, which can lead to false alarms. Consistent with this view, a recent lesion study suggests that the PHG is responsible for the elevated false positives for living distractors (high interference) than nonliving distractors (Kivisaari, Monsch, & Taylor, 2013). By linking the neural representations in the PHG with memory performance, our results could account for both true and false memories as observed in the current study and in the Kivisaari et al. (2013).

Using RSA on data from pictures representing living and nonliving objects, a recent study revealed a gradient of informational specificity: Whereas the early visual cortex represented image-based visual properties, and the posterior ventral stream represented the categorical information, the anterior part of the PHG (perirhinal cortex) uniquely represented object-specific semantic information and was engaged in the processing of semantically confusable objects (Clarke & Tyler, 2014). This highly conjunctive object-specific representation in the PHG provides an important insight regarding the mechanisms involved in the rejection of novel yet confusable items. The PHG showing semantic similarity in the current study was more posterior than that found in the previous study. Currently, it is not clear whether this difference could be due to the use of different stimuli (words vs. pictures). Importantly, the current study is limited in its practical spatial resolution due to the use of standard image acquisition resolution (3mm cubic) and the searchlight method (12mm sphere). Future studies

should use high-resolution imaging and individualized ROI analysis to examine the role of MTL sub-regions in semantic similarity and episodic memory.

Consistent with previous studies (Lipp et al., 2004; New et al., 2007; Pratt, Radulescu, Guo, & Abrams, 2010), we found that living words were processed faster in semantic judgment. At the neural level, living words showed higher activity levels than nonliving words in the dorsal attention network, which is posited to be involved in top-down attention process (Corbetta et al., 2008; Kim, 2011; Xue et al., 2013). However, neither the response time nor activity in the dorsal attention network was found to be predictive of memory performance, nor did they mediate the animacy effect in memory. Indeed, we only found a trend of the SME in this region. One reason for the lack of SME in this region may be the short delay between encoding and memory test (30 minutes). On the one hand, the short delay led to a high overall memory performance, which might have undermined the power to detect the SME. On the other hand, a previous study has suggest that the SME in the lateral prefrontal cortex increased with the delay between study and test, with stronger lateral prefrontal cortex activity supporting more durable memories (Q. Liu, Dong, Chen, & Xue, 2014). Future studies should increase the delay to see whether attention contributes to the animacy effect.

As mentioned earlier, stronger activation in the LHIP for living words than for nonliving words also mediated the animacy effect. According to a recent review (Brod, Werkle-Bergner, & Shing, 2013), the hippocampus plays a prominent role in linking prior knowledge to memory encoding. In particular, the hippocampus is suggested to integrate information to form a context for memory, which is then bound to new inputs to form episodic memories (Mizumori, 2013). Based on this assumption, higher activation in the LHIP during encoding could suggest greater context building and also better memory performance. In the current study, since living words were found to share more semantic features than nonliving words, the greater activations in the LHIP could reflect a stronger attempt to build semantic contexts for living words, as suggested by the context retrieval models (Howard & Kahana, 2002; Morton et al., 2013; Polyn, Norman, & Kahana, 2009). Specifically, during encoding, category-specific retrieval cues were

constructed to allow them to target items from certain categories during memory search. Using EEG to characterize category-specific oscillatory activity, Morton (2013) found that as multiple same-category items were studied sequentially, item-specific patterns decreased in strength whereas category-specific patterns increased in strength. The strength of category-specific patterns in the temporal lobe predicted items' subsequent memory performance.

The interplay between semantic memory and episodic memory is further supported by lesion studies. For example, patients with semantic dementia show impairment in episodic memory (Graham, Simons, Pratt, Patterson, & Hodges, 2000; Irish, Addis, Hodges, & Piguet, 2012; Maguire, Kumaran, Hassabis, & Kopelman, 2010). Notably, although Tulving's (1995) SPI model posits that the encoding is serial (i.e., perceptual information is input to semantic memory, and then episodic memory is based on the output from semantic memory), Graham's (2000) study found that semantic deficit could only impair episodic memory when the perceptual cues were not available. This finding suggests that perceptual information could provide direct input to the episodic memory system to aid memory. Future studies should examine whether the global match mechanism could operate on the perceptual features, by using the approach proposed in the current study.

In summary, our research found that nGPS reflected the nature of semantic organization, with living words showing a greater overlap in semantic features than nonliving words, which led to better memory performance for the former. This finding thus provides neural evidence to support the interaction between semantic memory and episodic memory. Nevertheless, the semantic similarity ratings used in the current study provide only one of many possible ways to describe semantic representations (Griffiths, Steyvers, & Tenenbaum, 2007; Zhao, Li, & Kohonen, 2011), future studies should examine whether alternative methods would provide a better account of the interaction between semantic memory and episodic memory. Furthermore, it is still unclear what is the "memory signal" driven by the neural global match (Norman, 2010). Future research

needs to understand how the brain works to obtain the global match information to aid memory performance.

### **Acknowledgments**

This work was partly sponsored by the National Natural Science Foundation of China (31130025), The 973 Program (2014CB846102), the 111 Project (B07008), and the NSFC Project (31221003).

ACCEPTED MANUSCRIPT

Figure 1. Experiment design and data analysis. (A) A slow event-related design (12s for each trial) was used to improve the accuracy in the estimation of single-trial responses. Each trial started with 1s fixation. Each word was presented for 3s. Participants were asked to judge whether the word represents a living or nonliving object. To prevent further encoding of the word, a series of Gabor images were presented during the 8s inter-trial-interval, and subjects were asked to judge the orientation of the Gabor images as quickly and accurately as possible. (B) Calculation of neural pattern similarity. Pearson correlations were performed for each participant for each stimulus to assess neural pattern similarity. Global pattern similarity (GPS) was calculated by averaging the z-transformed correlations between one item with all other items. This GPS could be further decomposed into within-category (WC) similarity and across-category (AC) similarity.

Figure 2. Behavioral results. (A) Memory performance. The left y-axis represents the percentage of hit (left two bars) or false alarm (FA, two bars in the middle), and the right y-axis represents the  $d'$ . (B) Reaction time and (C) accuracy during the semantic judgment task, as a function of subsequent memory, semantic category, and repetitions. Error bars represent within-subject error. R: recognized; F: forgotten; N: nonliving; L: living.  $**p < .01$ ,  $*p < .05$ .

Figure 3. Semantic similarity. (A) Similarity matrix of subjective ratings of words. Secondary pair-wise similarities were averaged across subjects for each item. Living words showed higher global similarity than did nonliving words. (B) Multidimensional scaling of pairwise similarities of words' semantic similarity. Distances between items in the plot depict dissimilarities between their respective semantic similarities in the semantic space. (C) Global semantic similarity as a function of semantic categories. Error bars represent the within-subject error.  $**p < .01$ .

Figure 4. The animacy effect in activation level. (A) The animacy effect in brain regions sensitive to SME. Focusing on the five clusters showing significant subsequent memory effects, including the LHIP, LaPHG, LOFC, LLOC and RFG (as shown on the activation map), mixed-effect models revealed significant animacy effects (greater activation for living words than for nonliving words) in the LHIP ( $p = .019$ , corrected) and marginally significant effect in the LaPHG ( $p = .09$ , corrected). Error bars represent the within-subject error. (B) Mediation analysis showed that the better memory performance for the living words was mediated by the mean activation level in LHIP. \*\*\* $p < .001$ , \* $p < .05$ , # $p < .1$ .

Figure 5. The animacy effect in neural global pattern similarity (nGPS). (A) The MTL and ATL mask (based on the Harvard-Oxford probabilistic anatomical atlas, thresholded at 25% probability; purple: hippocampus; green: parahippocampus gyrus; yellow: anterior temporal lobe; blue: MTG). (B) The subsequent memory effect in the left MTL based on small volume correction ( $z = 2.3$ ,  $p < 0.05$ ). The nGPS in the left hippocampus and the posterior segment of left parahippocampus were sensitive to memory performance. (C) Differences between living and nonliving words in nGPS. The error bars represent within-subject error. \* indicates significant results based on the mixed-effects model test (uncorrected). (D) Mediation analysis showed that better memory performance for living words was driven by higher nGPS in left posterior segment of the parahippocampus. \*\*\*  $p < .001$ , \*\*  $p < .01$ .

## References

- Barensse, M. D., Groen, I. I., Lee, A. C., Yeung, L. K., Brady, S. M., Gregori, M., . . . Henson, R. N. (2012). Intact memory for irrelevant information impairs perception in amnesia. *Neuron*, *75*(1), 157-167. doi: 10.1016/j.neuron.2012.05.014
- Bates, D., Maechler, M., Bolker, B. (2012). lme4: 0: Linear mixed-effects models using Eigen and syntax (2012). *R package version 0.9999-1/r1692*, 2012.
- Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage*, *20*(2), 1052-1063. doi: 10.1016/s1053-8119(03)00435-x
- Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain Lang*, *98*(2), 127-139. doi: 10.1016/j.bandl.2006.04.008
- Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J., & Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cereb Cortex*, *20*(11), 2728-2738. doi: 10.1093/cercor/bhq019
- Bonin, P., Gelin, M., & Bugajska, A. (2014). Animates are better remembered than inanimates: further evidence from word and picture stimuli. *Mem Cognit*, *42*(3), 370-382.
- Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. *Front Behav Neurosci*, *7*, 139. doi: 10.3389/fnbeh.2013.00139
- Bussey, T. J., & Saksida, L. M. (2005). Object memory and perception in the medial temporal lobe: an alternative approach. *Curr Opin Neurobiol*, *15*(6), 730-737. doi: 10.1016/j.conb.2005.10.014
- Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc Natl Acad Sci U S A*, *111*(40), 14565-14570. doi: 10.1073/pnas.1402594111
- Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Curr Opin Neurobiol*, *17*(2), 177-184. doi: 10.1016/j.conb.2007.03.005
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychon Bull Rev*, *3*(1), 37-60. doi: 10.3758/bf03210740
- Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *J Neurosci*, *34*(14), 4766-4775. doi: 10.1523/jneurosci.2828-13.2014
- Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., . . . Haxby, J. V. (2012). The Representation of Biological Classes in the Human Brain. *Journal of Neuroscience*, *32*(8), 2608-2618.
- Copland, D. A., de Zubicaray, G. I., McMahon, K., Wilson, S. J., Eastburn, M., & Chenery, H. J. (2003). Brain activity during automatic semantic priming revealed by event-related functional magnetic resonance imaging. *Neuroimage*, *20*(1), 302-310.
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306-324. doi: 10.1016/j.neuron.2008.04.017



- Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. *J Neurosci*, *26*(47), 12186-12197. doi: 10.1523/jneurosci.2818-06.2006
- Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J Exp Psychol Gen*, *132*(2), 163-201.
- Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: a comment on Dennis and Humphreys (2001). *Psychol Rev*, *111*(3), 800-807. doi: 10.1037/0033-295x.111.3.800
- Davis, T., Xue, G., Love, B. C., Preston, A. R., & Poldrack, R. A. (2014). Global Neural Pattern Similarity as a Common Basis for Categorization and Recognition Memory. *The Journal of Neuroscience*, *34*(22), 7472-7484.
- Devlin, J. T., Matthews, P. M., & Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J Cogn Neurosci*, *15*(1), 71-84. doi: 10.1162/089892903321107837
- Dixon, M. J., Bub, D. N., & Arguin, M. (1998). Semantic and visual determinants of face recognition in a prosopagnosic patient. *J Cogn Neurosci*, *10*(3), 362-376.
- Fairchild, A. J., Mackinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R2 effect-size measures for mediation analysis. *Behav Res Methods*, *41*(2), 486-498. doi: 10.3758/brm.41.2.486
- Farah, M. J., & Rabinowitz, C. (2003). Genetic and environmental influences on the organisation of semantic memory in the brain: is "living things" an innate category? *Cogn Neuropsychol*, *20*(3), 401-408. doi: 10.1080/02643290244000293
- Gale, T. M., & Laws, K. R. (2006). Category-specificity can emerge from bottom-up visual characteristics: evidence from a modular neural network. *Brain Cogn*, *61*(3), 269-279. doi: 10.1016/j.bandc.2006.02.001
- Garrard, P., Lambon Ralph, M. A., Watson, P. C., Powis, J., Patterson, K., & Hodges, J. R. (2001). Longitudinal profiles of semantic impairment for living and nonliving concepts in dementia of Alzheimer's type. *J Cogn Neurosci*, *13*(7), 892-909. doi: 10.1162/089892901753165818
- Gerlach, C., & Marques, J. F. (2014). Visual complexity exerts opposing effects on object categorization and identification. *Vis cogn*, *22*(6), 751-769. doi: 10.1080/13506285.2014.915908
- Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic memory. *Neuropsychologia*, *38*(3), 313-324.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychol Rev*, *114*(2), 211-244.
- Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage*, *57*(2), 482-494.
- Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *J Mem Lang*, *46*(1), 85-98. doi: Doi 10.1006/Jmla.2001.2798

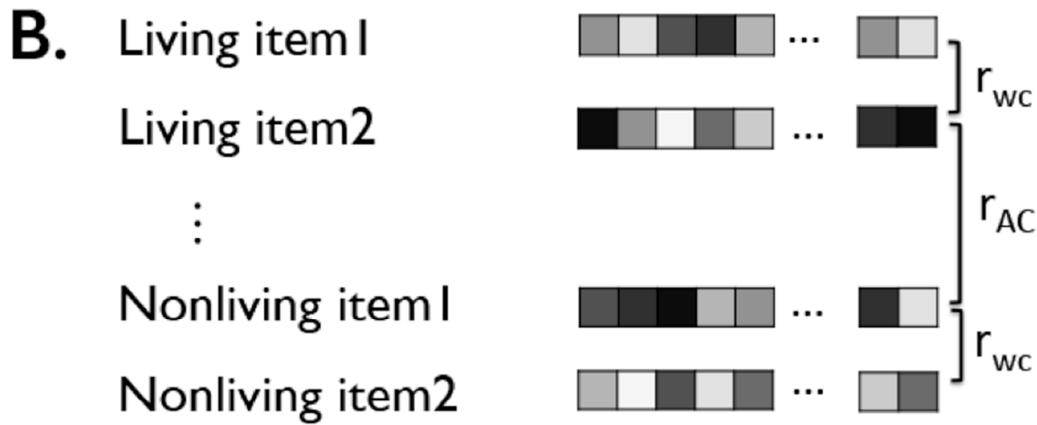
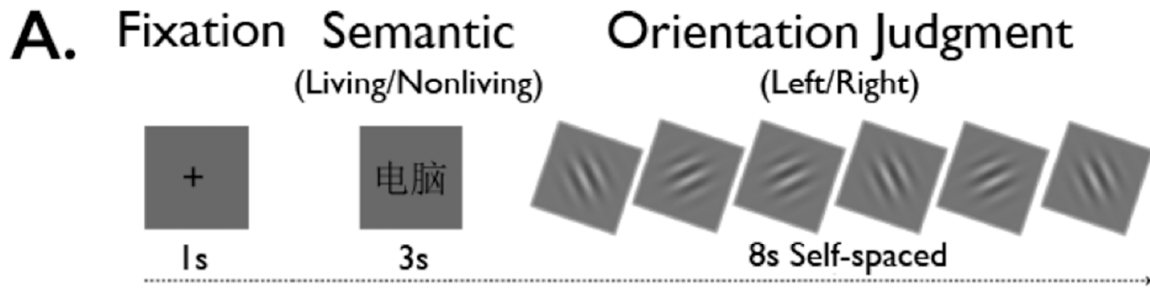
- Humphreys, G. W., & Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behav Brain Sci*, 24(3), 453-476; discussion 476-509.
- Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. *Brain*, 135(Pt 7), 2178-2191. doi: 10.1093/brain/aws119
- Islam, A., & Inkpen, D. (2006). *Second order co-occurrence PMI for determining the semantic similarity of words*. Paper presented at the Proceedings of the International Conference on Language Resources and Evaluation, Genoa, Italy.
- Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput Biol*, 10(11), e1003915. doi: 10.1371/journal.pcbi.1003915
- Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *Neuroimage*, 54(3), 2446-2461. doi: 10.1016/j.neuroimage.2010.09.045
- Kivisaari, S. L., Monsch, A. U., & Taylor, K. I. (2013). False positives to confusable objects predict medial temporal lobe atrophy. *Hippocampus*, 23(9), 832-841. doi: Doi 10.1002/Hipo.22137
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proc Natl Acad Sci U S A*, 103(10), 3863-3868. doi: 10.1073/pnas.0600244103
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci*, 17(8), 401-412. doi: 10.1016/j.tics.2013.06.007
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci*, 2, 4. doi: 10.3389/neuro.06.004.2008
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126-1141. doi: 10.1016/j.neuron.2008.10.043
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav Res Methods*, 44(4), 978-990. doi: 10.3758/s13428-012-0210-4
- Lag, T. (2005). Category-specific effects in object identification: what is "normal"? *Cortex*, 41(6), 833-841.
- Laiacina, M., Barbarotto, R., & Capitani, E. (2006). Human evolution and the brain representation of semantic knowledge: is there a role for sex differences? *Evolution and Human Behavior*, 27(2), 158-168. doi: 10.1016/j.evolhumbehav.2005.08.002
- Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia--what matters? *Neuropsychologia*, 36(8), 775-784.
- LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J Neurosci*, 33(13), 5466-5474. doi: 10.1523/JNEUROSCI.4293-12.2013

- Laws, K. R. (1999). Gender affects naming latencies for living and nonliving things: implications for familiarity. *Cortex*, 35(5), 729-733.
- Laws, K. R. (2000). Category-specific naming errors in normal subjects: the influence of evolution and experience. *Brain Lang*, 75(1), 123-133. doi: 10.1006/brln.2000.2348
- Laws, K. R. (2004). Sex differences in lexical size across semantic categories. *Personality and Individual Differences*, 36, 23-32.
- Laws, K. R., & Gale, T. M. (2002). Category-specific naming and the 'visual' characteristics of line drawn stimuli. *Cortex*, 38(1), 7-21.
- Laws, K. R., Gale, T. M., Frank, R., & Davey, N. (2002). Visual similarity is greater for line drawings of nonliving than living things: the importance of musical instruments and body parts. *Brain Cogn*, 48(2-3), 421-424.
- Laws, K. R., Leeson, V. C., & Gale, T. M. (2002). The effect of 'masking' on picture naming. *Cortex*, 38(2), 137-147.
- Laws, K. R., & Neve, C. (1999). A normal' category-specific advantage for naming living things. *Neuropsychologia*, 37(11), 1263-1269.
- Lipp, O. V., Derakshan, N., Waters, A. M., & Logies, S. (2004). Snakes and cats in the flower bed: fast detection is not specific to pictures of fear-relevant animals. *Emotion*, 4(3), 233-250. doi: 10.1037/1528-3542.4.3.233
- Liu, C. L., Hue, C. W., Chen, C. C., Chuang, K. H., Liang, K. C., Wang, Y. H., . . . Chen, J. H. (2006). Dissociated roles of the middle frontal gyri in the processing of Chinese characters. *Neuroreport*, 17(13), 1397-1401. doi: 10.1097/01.wnr.0000233090.00463.35
- Liu, Q., Dong, Q., Chen, C. S., & Xue, G. (2014). Neural processes during encoding support durable memory. *Neuroimage*, 88, 1-9. doi: Doi 10.1016/J.Neuroimage.2013.11.031
- Lu, Y., Wang, C., Chen, C., & Xue, G. (2015). Spatiotemporal neural pattern similarity supports episodic memory. *Curr Biol*, 25(6), 780-785. doi: 10.1016/j.cub.2015.01.055
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods*, 7(1), 83-104.
- Maguire, E. A., Kumaran, D., Hassabis, D., & Kopelman, M. D. (2010). Autobiographical memory in semantic dementia: a longitudinal fMRI study. *Neuropsychologia*, 48(1), 123-136. doi: 10.1016/j.neuropsychologia.2009.08.020
- Marques, J. F., Raposo, A., & Almeida, J. (2013). Structural processing and category-specific deficits. *Cortex*, 49(1), 266-275. doi: 10.1016/j.cortex.2011.10.006
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev*, 102(3), 419-457.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav Res Methods*, 37(4), 547-559.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *J Exp Psychol Gen*, 126(2), 99-130.

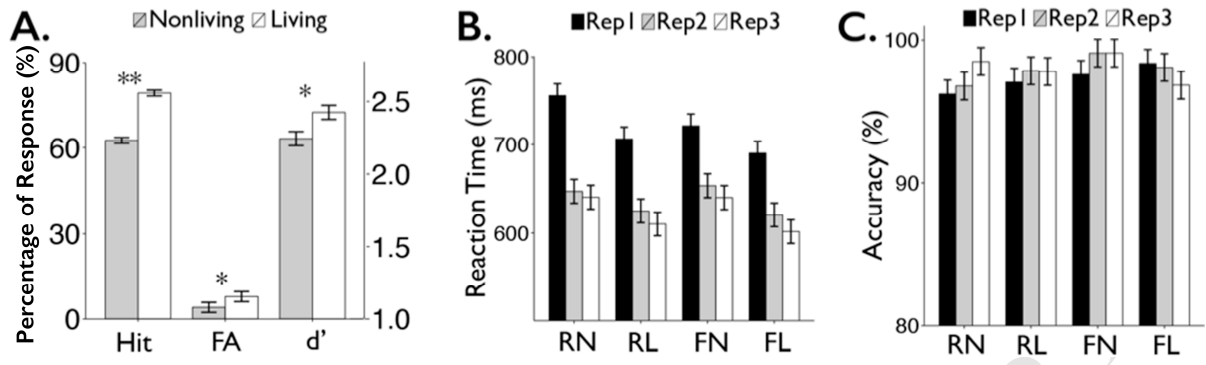
- McTighe, S. M., Cowell, R. A., Winters, B. D., Bussey, T. J., & Saksida, L. M. (2010). Paradoxical False Memory for Objects After Brain Damage. *Science*, *330*(6009), 1408-1410. doi: Doi 10.1126/Science.1194780
- Mizumori, S. J. (2013). Context prediction analysis and episodic memory. *Front Behav Neurosci*, *7*, 132. doi: 10.3389/fnbeh.2013.00132
- Morton, N. W., Kahana, M. J., Rosenberg, E. A., Baltuch, G. H., Litt, B., Sharan, A. D., . . . Polyn, S. M. (2013). Category-specific neural oscillations predict recall organization during memory search. *Cereb Cortex*, *23*(10), 2407-2422. doi: 10.1093/cercor/bhs229
- Mumford, J. A., & Poldrack, R. A. (2007). Modeling group fMRI data. *Soc Cogn Affect Neurosci*, *2*(3), 251-257. doi: 10.1093/scan/nsm019
- Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, *59*(3), 2636-2643. doi: 10.1016/j.neuroimage.2011.08.076
- Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Front Psychol*, *4*, 128. doi: 10.3389/fpsyg.2013.00128
- Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: the mnemonic value of animacy. *Psychol Sci*, *24*(10), 2099-2105. doi: 10.1177/0956797613480803
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc Natl Acad Sci U S A*, *104*(42), 16598-16603. doi: 10.1073/pnas.0703913104
- Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: revisiting the complementary learning systems model. *Hippocampus*, *20*(11), 1217-1227. doi: 10.1002/hipo.20855
- O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketzer, N. (2014). Complementary learning systems. *Cogn Sci*, *38*(6), 1229-1248. doi: 10.1111/j.1551-6709.2011.01214.x
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. *Psychol Rev*, *116*(1), 129-156. doi: Doi 10.1037/A0014420
- Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R. A. (2010). It's alive! animate motion captures visual attention. *Psychol Sci*, *21*(11), 1724-1730. doi: 10.1177/0956797610387440
- Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychol Bull*, *127*(2), 209-228.
- Rostad, K., Yott, J., & Poulin-Dubois, D. (2012). Development of categorization in infancy: advancing forward to the animate/inanimate level. *Infant Behav Dev*, *35*(3), 584-595. doi: 10.1016/j.infbeh.2012.05.005
- Russo, R., Mammarella, N., & Avons, S. E. (2002). Toward a unified account of spacing effects in explicit cued-memory tasks. *J Exp Psychol Learn Mem Cogn*, *28*(5), 819-829.
- Sartori, G., Gnoato, F., Mariani, I., Prioni, S., & Lombardi, L. (2007). Semantic relevance, domain specificity and the sensory/functional theory of category-

- specificity. *Neuropsychologia*, 45(5), 966-976. doi: 10.1016/j.neuropsychologia.2006.08.028
- Sartori, G., & Lombardi, L. (2004). Semantic relevance and semantic disorders. *J Cogn Neurosci*, 16(3), 439-452. doi: 10.1162/089892904322926773
- Sartori, G., Lombardi, L., & Mattiuzzi, L. (2005). Semantic relevance best predicts normal and abnormal name retrieval. *Neuropsychologia*, 43(5), 754-770. doi: 10.1016/j.neuropsychologia.2004.08.001
- Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *J Exp Psychol Learn Mem Cogn*, 21(2), 267-287.
- Silveri, M. C., Cappa, A., Mariotti, P., & Puopolo, M. (2002). Naming in patients with Alzheimer's disease: influence of age of acquisition and categorical effects. *J Clin Exp Neuropsychol*, 24(6), 755-764. doi: 10.1076/jcen.24.6.755.8407
- Su, L., Zulfiqar, I., Jamshed, F., Fonteneau, E., & Marslen-Wilson, W. (2014). Mapping tonotopic organization in human temporal cortex: representational similarity analysis in EMEG source space. *Front Neurosci*, 8, 368. doi: 10.3389/fnins.2014.00368
- Tofighi, D., & MacKinnon, D. P. (2011). RMediation: an R package for mediation analysis confidence intervals. *Behav Res Methods*, 43(3), 692-700. doi: 10.3758/s13428-011-0076-x
- Turk-Browne, N. B., Golomb, J. D., & Chun, M. M. (2013). Complementary attentional components of successful memory encoding. *Neuroimage*, 66, 553-562. doi: 10.1016/j.neuroimage.2012.10.053
- Turnbull, O. H. t., & Laws, K. R. (2000). Loss of stored knowledge of object structure: implications for "category-specific" deficits. *Cogn Neuropsychol*, 17(4), 365-389. doi: 10.1080/026432900380445
- Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiol Learn Mem*, 91(2), 139-154. doi: 10.1016/j.nlm.2008.10.011
- VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N., & Blunt, J. R. (2013). Adaptive memory: animacy processing produces mnemonic advantages. *Exp Psychol*, 60(3), 172-178. doi: 10.1027/1618-3169/a000186
- VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N., & Cogdill, M. (2014). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, 1-7. doi: 10.1080/09658211.2014.916304
- VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: animacy processing produces mnemonic advantages. *Exp Psychol*, 60(3), 172.
- Ward, E. J., Chun, M. M., & Kuhl, B. A. (2013). Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *J Neurosci*, 33(37), 14749-14757. doi: 10.1523/JNEUROSCI.4889-12.2013
- Woolrich, M. W. (2008). Robust group analysis using outlier inference. *Neuroimage*, 41(2), 286-301. doi: 10.1016/j.neuroimage.2008.02.042
- Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian

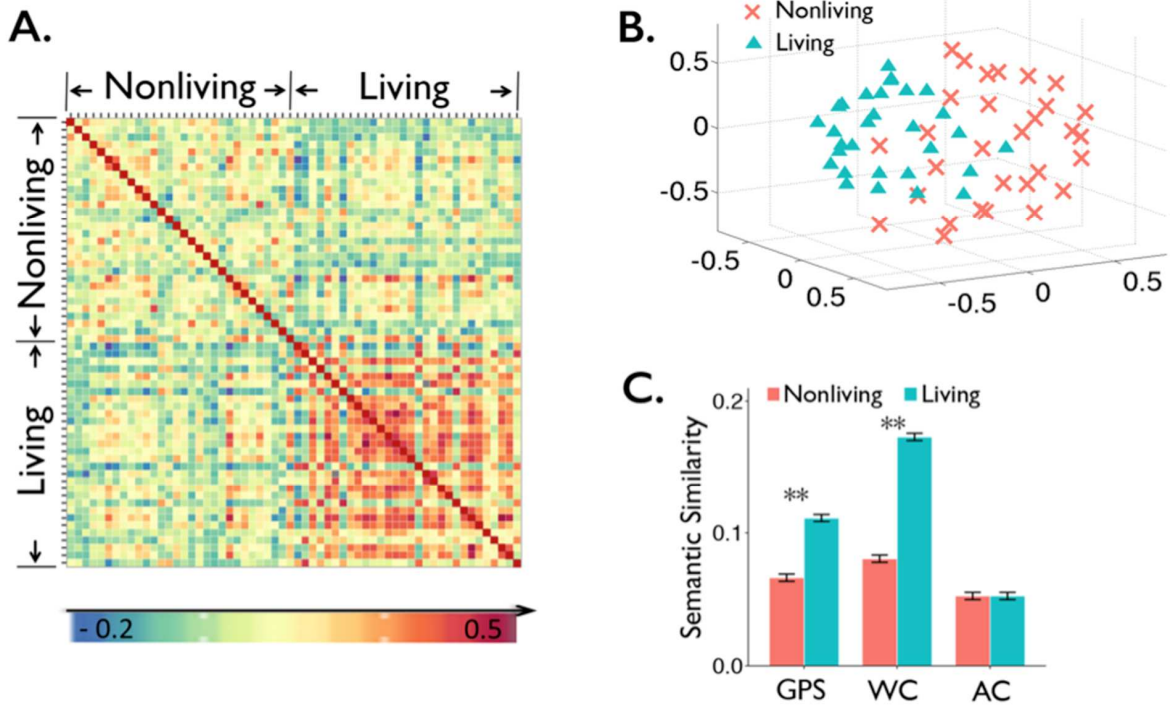
- inference. *Neuroimage*, 21(4), 1732-1747. doi: 10.1016/j.neuroimage.2003.12.023
- Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, 330(6000), 97-101. doi: 10.1126/science.1193125
- Xue, G., Dong, Q., Chen, C., Lu, Z. L., Mumford, J. A., & Poldrack, R. A. (2013). Complementary role of frontoparietal activity and cortical pattern similarity in successful episodic memory encoding. *Cereb Cortex*, 23(7), 1562-1571. doi: 10.1093/cercor/bhs143
- Zannino, G. D., Perri, R., Pasqualetti, P., Caltagirone, C., & Carlesimo, G. A. (2006a). Analysis of the semantic representations of living and nonliving concepts: A normative study. *Cogn Neuropsychol*, 23(4), 515-540. doi: 10.1080/02643290542000067
- Zannino, G. D., Perri, R., Pasqualetti, P., Caltagirone, C., & Carlesimo, G. A. (2006b). (Category-specific) semantic deficit in Alzheimer's patients: the role of semantic distance. *Neuropsychologia*, 44(1), 52-61. doi: 10.1016/j.neuropsychologia.2005.04.008
- Zannino, G. D., Perri, R., Pasqualetti, P., Di Paola, M., Caltagirone, C., & Carlesimo, G. A. (2006). The role of semantic distance in category-specific impairments for living things: evidence from a case of semantic dementia. *Neuropsychologia*, 44(7), 1017-1028. doi: 10.1016/j.neuropsychologia.2005.11.006
- Zhao, X. W., Li, P., & Kohonen, T. (2011). Contextual self-organizing map: software for constructing semantic representations. *Behav Res Methods*, 43(1), 77-88. doi: Doi 10.3758/S13428-010-0042-Z

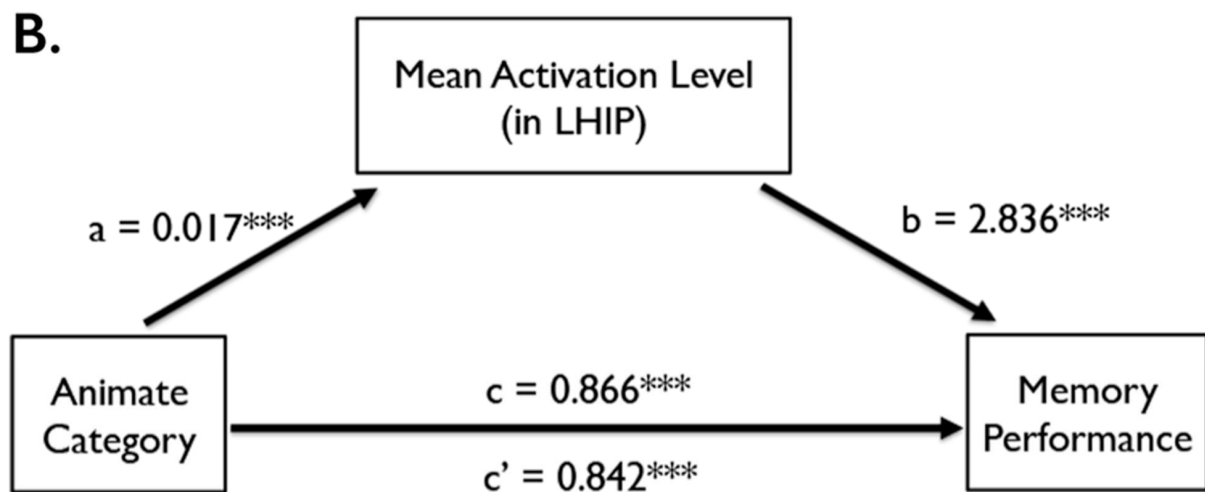
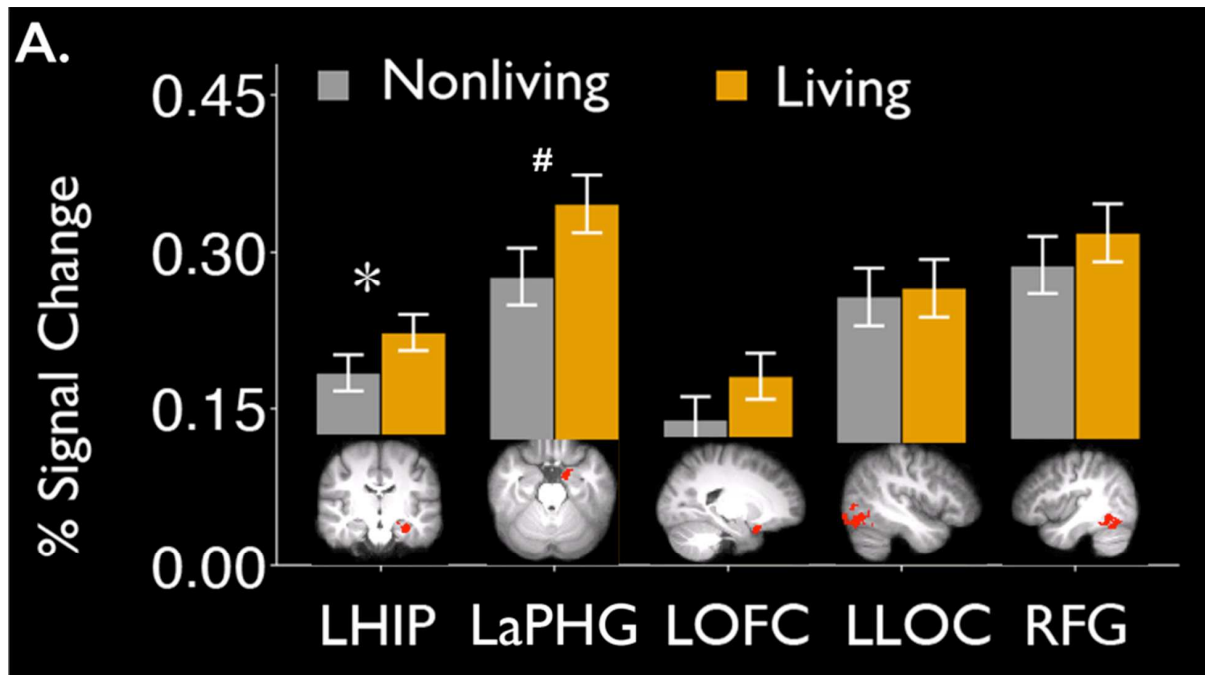


$$GPS = (\sum r_{WC} + \sum r_{AC}) / (N_{WC} + N_{AC})$$

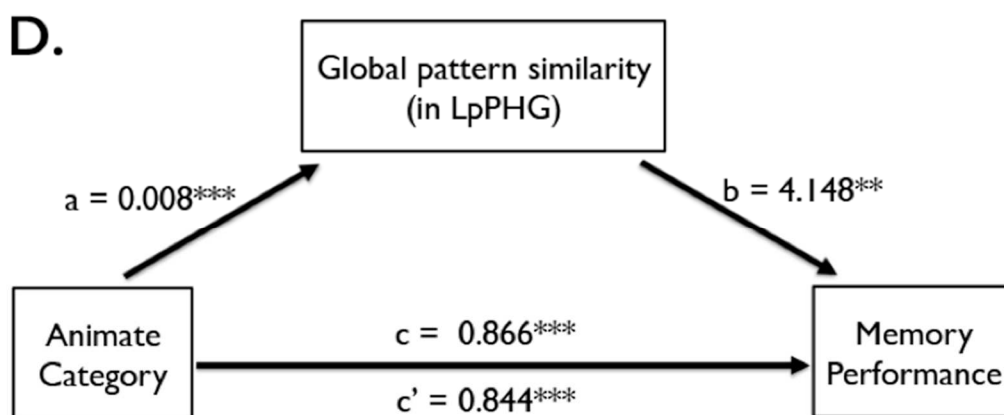
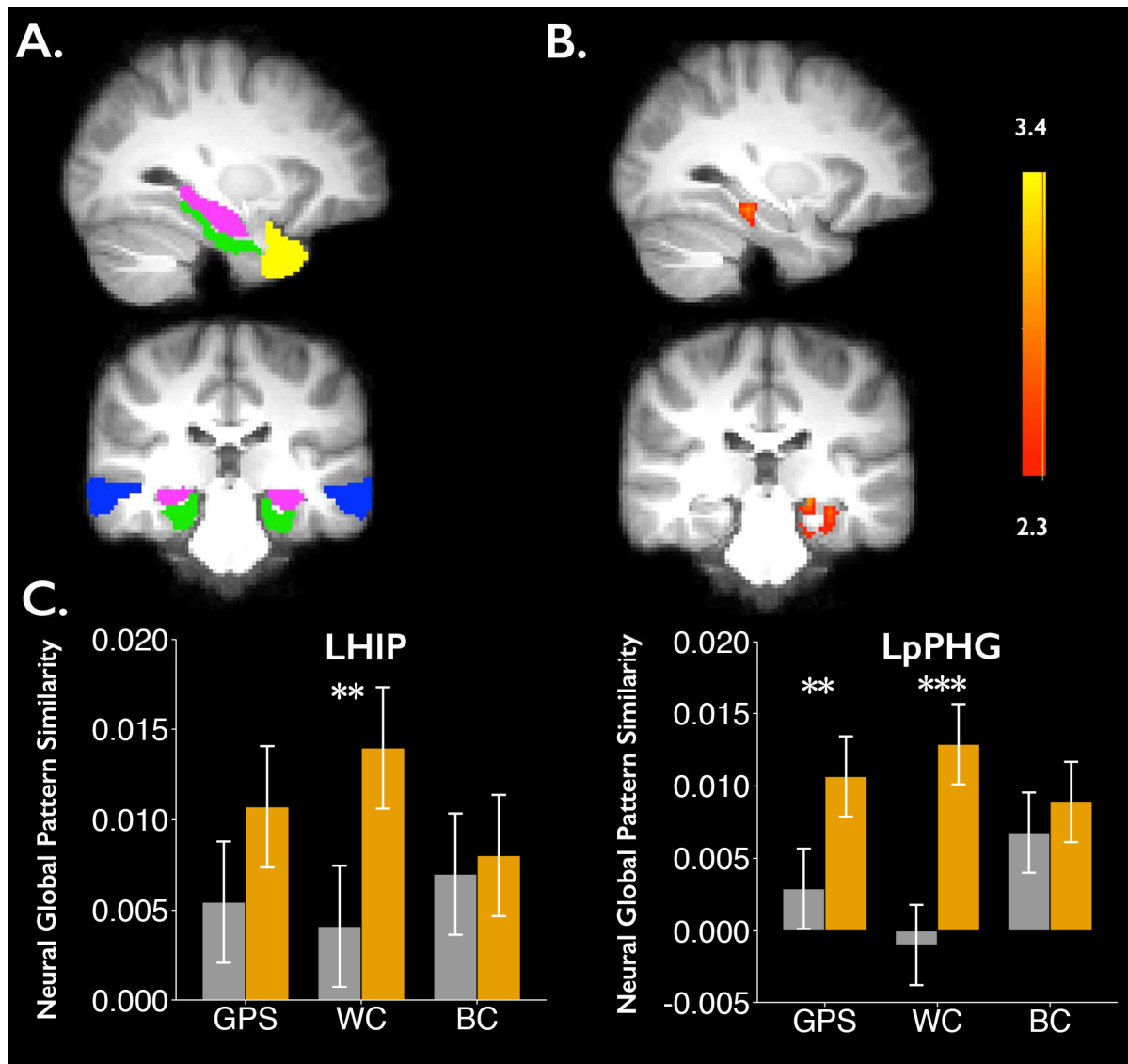








Indirect effect = 0.045 (SE = 0.022), 95% CI is [0.008, 0.092]



Indirect effect = 0.028 (SE = 0.016), 95% CI is [0.003, 0.065]