Adjudicating competing theories of semantic representation with 7T-fMRI and multivariate decoding

Saskia L. Frisby[1], Christopher R. Cox[2], Ajay D. Halai[1], Matthew A. Lambon Ralph[1], & Timothy T. Rogers[3]

[1] MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

[2] Department of Psychology, Louisiana State University, Baton Rouge, USA

[3] Department of Psychology, University of Wisconsin–Madison, Madison, USA

Abstract: Theories of semantic representation in the brain are numerous and contradictory and are supported by an equally contradictory literature of multivariate neuroimaging studies. Equipped with two pioneering tools – (1) a 7T-fMRI acquisition sequence capable of capturing signal in crucial ventral temporal regions, and (2) four innovative decoding methods that are designed to probe specific hypotheses about the nature of the semantic code – we resolve this confusion. Evidence best supports the hypothesis that the ventral anterior temporal lobe represents complex, multidimensional semantic information via an interdependent, dynamic code similar to that observed in neural networks. Similar representations – highly variable across individuals – were discovered in posterior temporal and occipitotemporal cortex. In summary, this work adjudicates competing theories of semantic cognition, thereby reconciling the discrepancy in results between multivariate imaging and other sources of evidence – neuropsyschology, noninvasive brain stimulation, intracranial electrophysiology, and neural network modelling.

Keywords: semantic representation, 7T-fMRI, decoding, multivariate pattern analysis

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript arising from this work.

1. Introduction

Semantic cognition is our ability to store general knowledge and to use this knowledge to interpret the meaning of language, objects, and events. Influential theories of how the brain might encode semantic information are numerous and contradictory[1–9] and are supported by an equally contradictory literature of multivariate neuroimaging studies[10–18]. This confusion arises for at least two reasons – (1) magnetic field inhomogeneity[19–26] that compromises signal quality in crucial ventral temporal regions and (2) analytic assumptions, embedded in multivariate methods, that restrict the kinds of hypotheses that each method can test[27]. This study surmounted these obstacles by (1) utilising a novel 7 Tesla functional magnetic resonance imaging (7T-fMRI) acquisition protocol capable of recovering signal in regions prone to imaging artefacts[26] and (2) employing four innovative decoding methods[11,28–31] that are designed to directly adjudicate different hypotheses about the nature of the semantic code. Previous theoretical work described what semantic representations might possibly be like[27] – here, for the first time, we resolve those outstanding questions empirically.

Despite their many differences, multiple theories of semantic cognition agree that the ventral anterior temporal lobes (vATLs) are important (cf.[7,8]). This claim is supported by multiple sources of evidence – in semantic dementia, atrophy of the vATL causes multimodal semantic impairment but spares other cognitive abilities[32–35] and a similar, though milder and transient, effect is observed following transcranial magnetic stimulation (TMS) to the vATL[23,36–39]. However, theorists disagree on the vATL's precise role. Damasio and colleagues[3] propose that modality-specific semantic information (an object's visual appearance, name, associated motor action, etc.) is encoded in separate modality-specific regions; convergence zones – one for each concept – trigger and synchronise activity within these areas during representation. The convergence zones are organised, by category, into convergence regions, with a convergence region for unique entities (such as individuals) in the temporal pole and a convergence region for animals just behind[4]. By contrast, Binder and colleagues[2] propose that the vATL functions, not simply as a pointer towards modality-specific cortex, but as one of multiple semantic "hubs" that computes complex semantic representations from independent features (e.g. *has eyes*, *has gills*) represented elsewhere[1,2,40]. More similar concepts share more features, accounting for graded semantic

similarity (e.g. a robin and a wren are more similar to each other than to an ostrich). A third theory, the hub-and-spoke model[6,9,41], is motivated by the finding that neuro-computational models with a single hub can represent transmodal semantic structure (whereas other architectures fail[9,42,43]) and proposes that the vATL is the brain's *only* semantic hub. Like the deep layers of a computational model, activity within the vATL does not correspond to interpretable features – rather, after initial dynamic and nonlinear change, activity patterns adopt steady states that are gradedly more similar for more similar concepts.[44]

Notwithstanding its theoretical importance, neuroimaging studies of semantic tasks rarely highlight the vATL. One possible reason is the proximity of the vATL to the air-filled sinuses – this causes magnetic field inhomogeneity, and consequently signal dropout and distortions. Univariate fMRI studies that use distortion correction, such as spin-echo or multi-echo acquisition, do detect semantic effects in the vATL[19–26]. A second possible reason concerns analysis. Multivariate methods – which analyse patterns of activity over multiple neural units (voxels, sensors, or electrodes) to characterise what information those patterns represent – each encapsulate hypotheses that constrain the kind of neural code that they can detect and these assumptions can lead to drastically different results[27]. For example, when decoding semantic information from voltages recorded from electrocorticography (ECoG) surface grid electrodes, representational similarity analysis[45] revealed a simple binary animacy (living/non-living) distinction in the ATL[46] whereas methods based on regularised regression revealed a complex, multidimensional semantic structure[28].

Each contemporary theory of semantic cognition also implicates regions outside the vATL. Several theories agree that modality-specific regions are important because they encode modality-specific semantic information about a concept[2,3,6–8]. For example, neuroimaging studies highlight the inferior parietal cortex when processing tools[47]; lesions to inferior parietal regions can cause category-specific semantic impairments for tools[48]; and similar but transient impairments can be achieved with TMS[39]. Additionally, neuroimaging studies report activity in posterior medial temporal cortex when processing tools and in posterior lateral temporal cortex when processing animals[7,49]. However, theorists interpret this evidence differently – some assert that this indicates representation of interpretable features belonging to animals and tools; others note that the same results emerge in computational models with uninterpretable activations due to differential connectivity patterns of posterior temporal subregions[50]. Theories that propose the existence of multiple

hubs assign a central role to the angular gyrus, which is highlighted by imaging studies of semantic tasks[1,51] and which, when damaged, produces a pattern of impairment that appears to be mostly semantic[52] (cf.[53,54]). Finally, some theories even propose that semantic information is radically dispersed across most or all of the cortex[5,14,15]. Each of these opposing theories is supported to some extent by multivariate neuroimaging studies[10,12–15,28,44]. Although regions outside the ventral anterior temporal and orbitofrontal cortices are not affected by magnetic field inhomogeneity, methodological choices (which incorporate theoretical assumptions) may contribute to this variability.

In summary, contemporary theories of semantic cognition make contradictory claims. These claims concern not only *where* semantic information is represented, but also *how* it is represented – for example, Binder and colleauges[2] argue that semantic representations are constructed through linear combination of interpretable features, whereas the hub-and-spoke model[6,9,41] proposes that semantic features are uninterpretable and interdependent. Most multivariate neuroimaging studies test the predictions of a single theory; this work is distinctive because we directly tested the predictions of competing theories against one another. Utilising a clear theoretical framework[27], we tested whether semantic representations are:

1. *Category* representations (one discrete, independent unit per concept), *feature-based* representations (multiple independently-interpretable features), or *vector-space* representations (patterns across units that are not independently interpretable)
2. *Homogeneous* (units all adopt the same activation state – i.e., all increase in activity or all decrease – when representing the same concept) or *heterogeneous* (units adopt different activation states for the same concept)
3. *Consistent* (units encode semantic information in the same way across individuals) or *inconsistent* (units exhibit different semantic codes across individuals)
4. *Contiguous* (units are spatially adjacent) or *dispersed* (units are spatially distributed across the cortex) – representations can be contiguous or dispersed within or across individuals

We also assessed the compatibility of our results with the hub-and-spoke model's claim that representations in the vATL are *dynamic*, meaning that the information carried by a

128 particular neural unit changes over time – for example, a unit may signify animacy with
129 increased activity at one timepoint and with decreased activity at the next[44].
130 We overcame two obstacles that have hindered previous attempts to characterise
131 semantic representations. First, to address magnetic field inhomogeneity affecting ventral
132 temporal regions, we employed a novel multi-echo, multiband acquisition sequence
133 optimised for recovering signal in the vATL without sacrificing signal quality elsewhere[26].
134 Second, we leverage the inseparability of multivariate analysis and theory to our advantage
135 – we employ multivariate methods that make opposing theoretical assumptions and then
136 contrast the results obtained with each method. First, we decoded animacy using binary
137 logistic regression classifiers. These were used in combination with LASSO (L1)
138 regularisation[55], which assumes that the neural code is sparse (only a few of the features
139 under consideration encode the target structure), and with sparse-overlapping-sets LASSO
140 (SOSLASSO) regularisation[11,30,31], which additionally assumes that important features are
141 loosely spatially clustered within and across participants. Second, we used Representational
142 Similarity Learning (RSL)[28,29] to decode graded, multidimensional semantic similarity
143 structure. In RSL, a similarity structure is modelled (for example, using feature norms[27,56]),
144 then decomposed into multiple independent dimensions. Linear regression models then
145 learn to predict the coordinates of each stimulus on each dimension; significant correlations
146 between true and predicted coordinates indicate that the modelled similarity structure is a
147 good representation of the brain's. With RSL we used LASSO regularisation and group-
148 ordered-weighted LASSO (grOWL) regularisation[28,29], which assumes that the neural code is
149 sparse, that important features are correlated in their activity, and that features are likely to
150 encode multiple target dimensions at once (an interdependent, vector-space
151 representation).
152 We applied regularised logistic regression and RSL within a vATL region of interest
153 (ROI) and at the whole-brain level with the aim of adjudicating contemporary theories about
154 the nature of semantic representation.
155
156 **2. Results**
157
158 2.1. Characterising semantic representations in the ventral anterior temporal lobe
159 (vATL)

27 healthy participants completed the study successfully. None were excluded because of excessive head motion, as defined by two metrics – framewise displacement[57] and DVARS (which indexes the rate of change of BOLD signal across the entire brain between each pair of successive frames of data)[58]. The participants named 100 line drawings of animals and inanimate objects that were used previously for decoding [28,44] and that were matched for age of acquisition, concreteness, word frequency and familiarity. The participants named the stimuli with high accuracy (mean = 96.85%, standard deviation 2.71%). Three participants had at least one run with accuracy less than two standard deviations below the group mean. However, since none of these participants had accuracy less than 88% on any run, we decided to include them.

Multiple theories propose that the vATL plays an important role in semantic cognition[2,3,6,9,41], although its precise role is debated – a convergence region[3] or a hub, combining independent[2,7,8] or interdependent features[6,9,41]. Therefore, we aimed first to decode semantic information from the vATL and next to characterise the nature of semantic representations there. We constructed bilateral ventral temporal ROIs based on ECoG electrode coverage in a previous decoding study[59]. We also created "half-ROIs" by bisecting the ROIs between their anteriormost and posteriormost points. This was primarily to facilitate tests for dynamic coding, which was previously observed more distinctly in anterior than posterior temporal regions in ECoG data[44], but for completeness we ran all decoding analyses on all ROIs and half-ROIs. The ROIs were backprojected into each participant's native space and were masked with a participant-specific grey matter mask.

All models were trained on contrast beta values (one per stimulus, averaged over repeated presentations of the same stimulus) within each ROI and were assessed using ten-fold cross-validation.

2.1.1. Is semantic structure present in the vATL?

2.1.1.1. Animacy

First, we attempted to decode animacy from the vATL using logistic regression. Figure 1A shows the results. All classifiers trained on all ROIs performed better than chance.

Classification accuracy was very high (hold-out accuracy ≅ 0.95) for decoders trained on the whole ROI or on the posterior half. This result confirmed that animacy was represented within our ROI.

2.1.1.2. Multidimensional semantic structure

Second, we attempted to decode multidimensional semantic structure from the vATL using RSL. We modelled a target semantic similarity structure using feature norms[56] and used singular value decomposition to decompose the structure into three dimensions that explained 89.5% of the variance. Figure 2 shows the coordinates of all 100 stimuli on each dimension. We then trained RSL models to predict the coordinates of each stimulus on all dimensions.

Figure 1B shows the mean hold-out correlation between predicted and true coordinates on each target semantic dimension. All models showed very good decoding of dimension 1 of the target semantic space in all ROIs. In the right hemisphere, all models produced significant correlations on at least one other dimension that, though small, were statistically significant as assessed via permutation testing (see Methods; dimension 3 for LASSO, and both dimension 2 and dimension 3 for grOWL). In the left hemisphere, dimensions 2 and 3 were not reliably decoded in any ROI. These results confirmed the presence of multidimensional semantic structure in the right ROI.

2.1.1.3. Graded semantic structure

As shown in Figure 2, animate stimuli usually take low values on dimension 1 and inanimate stimuli usually take high values. Therefore, a model that predicted the same low value for all animate stimuli and the same high value for all inanimate stimuli could generate predicted coordinates that correlate reliably with the target coordinates, yet the model would not be detecting graded similarity structure. However, significant correlations *within* animate and inanimate domains (i.e. when the model is evaluated only on the animate or only on the inanimate stimuli) arise only if semantic structure is truly graded.[28]

Figure 3 shows correlations calculated for animate and inanimate stimuli separately. For animate stimuli, results depended on the regularisation penalty - whereas models with

224 LASSO exhibited reliable decoding of only one dimension, models with grOWL discovered
225 within-domain structure for dimensions 1 and 2. For inanimate stimuli, although correlation
226 coefficients were generally smaller, multiple dimensions were decoded. Whereas these
227 patterns were somewhat variable for models regularised with LASSO, models with grOWL
228 systematically showed significant decoding of dimensions 1 and 3 in both hemispheres.
229 These results confirmed that graded semantic structure was represented within our ROI.
230 When the two domains were considered separately, multidimensional semantic
231 structure was observed in the left temporal lobe for both, despite the fact that correlations
232 on only one dimension were observed there when all stimuli were considered together
233 (Figure 1). Note that dimension 2 distinguishes the animate stimuli, but not the inanimate
234 stimuli, and the reverse is true for dimension 3 (Figure 2); significant correlations within-
235 domain usually emerged on dimension 2 for animate stimuli and on dimension 3 for
236 inanimate stimuli.
237
238 2.1.2. What are semantic representations in the vATL like?
239
240 2.1.2.1. Categories, features, or vector-spaces?
241
242 Damasio and colleagues[3] propose that the vATL contains category-specific convergence
243 regions for animals and unique entities. This theory would predict successful decoding of
244 animacy from our ROI – it would be active for animate stimuli and not for inanimate stimuli.
245 However, because convergence zones simply point to representations elsewhere – they do
246 not themselves express semantic information – the vATL would not be predicted to express
247 multiple semantic dimensions. Figures 1 and 3 show that our results are at odds with this
248 prediction.
249 Of the theories that do predict multidimensional information representation in the
250 vATL, some propose that its role is to combine independently-interpretable features[2] while
251 others propose that it encodes a vector-space with interdependent, uninterpretable
252 dimensions[6,9,41]. These possibilities can be adjudicated by comparing RSL results with
253 different regularisation penalties. grOWL assumes that features are likely to encode multiple
254 target dimensions at once whereas LASSO does not make that assumption. Therefore,

255 better performance with grOWL than LASSO supports the notion of interdependent, vector-
256 space representation.
257 Figure 1B shows that RSL models with grOWL regularisation produced significant
258 correlations on all three dimensions, whereas models with LASSO regularisation produced
259 significant correlations on only two. Figure 3 shows that, for animate stimuli, models with
260 LASSO exhibited reliable decoding of only one dimension; models with grOWL could decode
261 dimensions 1 and 2. Within the inanimate domain, models with grOWL systematically
262 showed significant decoding of dimensions 1 and 3 whereas LASSO exhibited less consistent
263 results. Although grOWL produced larger correlations than LASSO (permutation $p < 0.05$) in
264 only a few cases, the reverse was never true. These results suggest that the semantic code
265 within the vATL is an uninterpretable, vector-space code.
266
267 2.1.2.2. Contiguous or dispersed?
268
269 SOSLASSO assumes that features are loosely spatially clustered within and across
270 participants, but LASSO does not make that assumption. Figure 1A shows that, in all ROIs
271 apart from the right anterior half-ROI, SOSLASSO produced significantly higher classification
272 accuracy than LASSO (permutation $p < 0.05$), suggesting that semantic representations
273 within the vATL are loosely contiguous within and across individuals.
274
275 2.1.2.3. Static or dynamic?
276
277 The hub-and-spoke model[6,9,41] claims that activity patterns within the vATL undergo
278 dynamic and nonlinear change during stimulus processing – for example, a voxel may signify
279 animacy with increased activity at one timepoint and with decreased activity at the next. A
280 previous ECoG study, focused on decoding animacy, observed these patterns more distinctly
281 in very anterior than in more posterior temporal regions[44]. The temporal resolution of fMRI
282 is far lower than that of ECoG and the resulting spatial blurring may therefore obscure a
283 temporally intricate semantic code. We would therefore expect poorer decoding
284 performance in the anterior than the posterior half of the ROI. Figure 2A shows that
285 classification accuracy in the anterior half-ROI was significantly worse than accuracy in the

posterior half-ROI (permutation $p < 0.05$). This fMRI result does not *establish* that the semantic code is dynamic, but the results are *compatible* with this claim.

Similarly, Figure 1B shows worse decoding in the anterior half-ROI than the posterior half ($r \cong 0.4$ vs. $r \cong 0.8$). However, dynamic representation of multidimensional semantics has not yet been demonstrated with time-resolved imaging methods. We therefore chose not to formally test the differences between half-ROIs or to interpret differences as evidence for dynamic coding of multidimensional semantic structure.

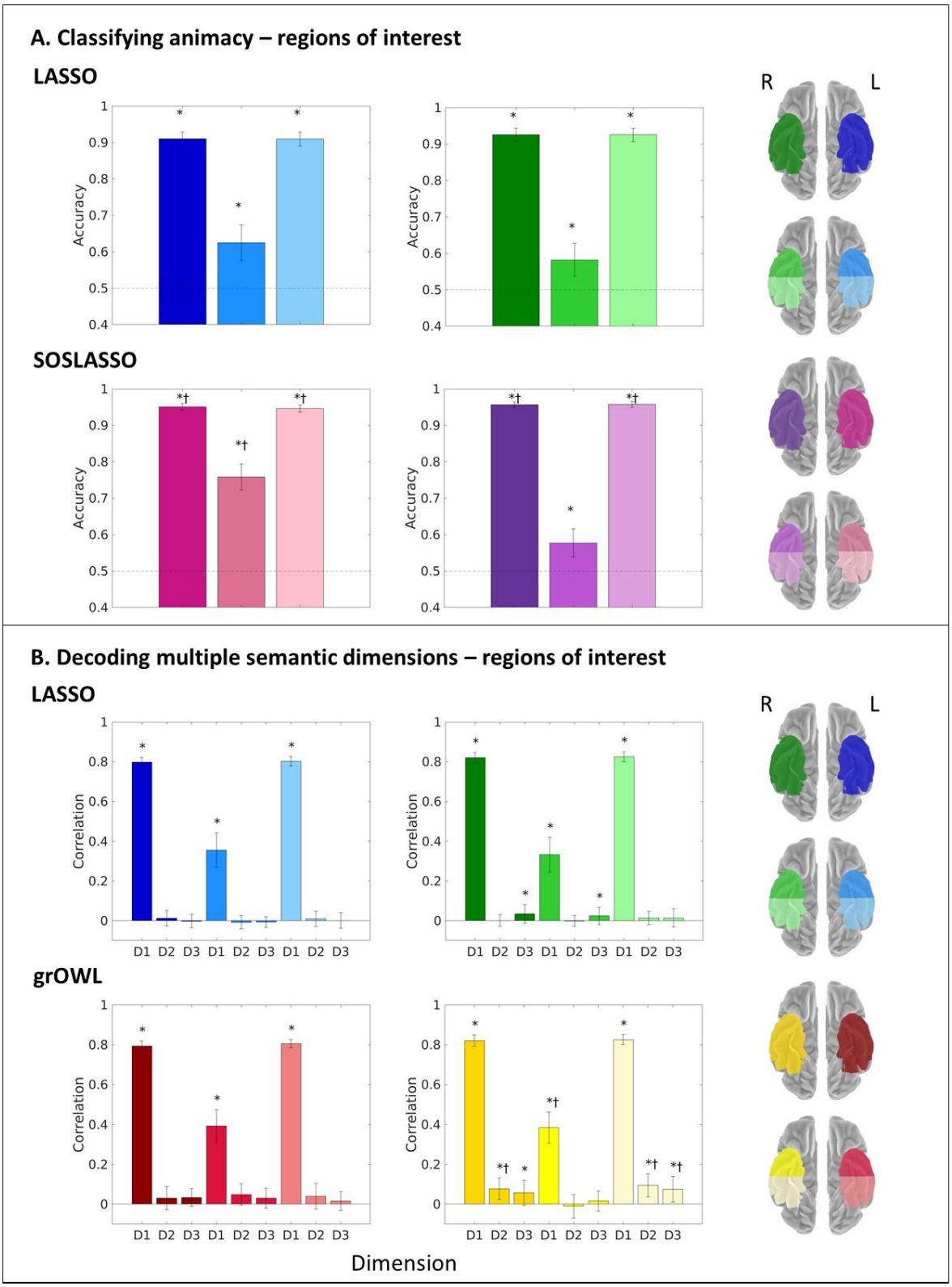

Figure 1: Decoding in regions of interest. Surface plots (right) indicate which bar corresponds to which ROI or half-ROI. (A) Mean and 95% confidence interval of the hold-out accuracy for regularised logistic regression classifiers trained on contrast beta values to discriminate animate from inanimate stimuli. Classifiers are trained using LASSO regularisation on the left ROI and half-ROIs (upper left, shades of blue), using LASSO regularisation on the right ROI and half-ROIs (upper right, shades of green), using SOSLASSO

regularisation on the left ROI and half-ROIs (lower left, shades of pink), or using SOSLASSO regularisation on the right ROI and half-ROIs (lower right, shades of purple). Stars (*) indicate a significant difference between classifier accuracy and chance (as determined by permutation test, probabilities adjusted to control the false-discovery rate at α = 0.05). Daggers (†) indicate a significant difference between classifier accuracy using that regularisation penalty and accuracy in the same ROI using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at α = 0.05). (B) Mean and 95% confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Models are trained using LASSO regularisation on the left ROI and half-ROIs (upper left, shades of blue), using LASSO regularisation on the right ROI and half-ROIs (upper right, shades of green), using grOWL regularisation on the left ROI and half-ROIs (lower left, shades of red), or using grOWL regularisation on the right ROI and half-ROIs (lower right, shades of yellow). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at α = 0.05). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension in the same ROI using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at α = 0.05).

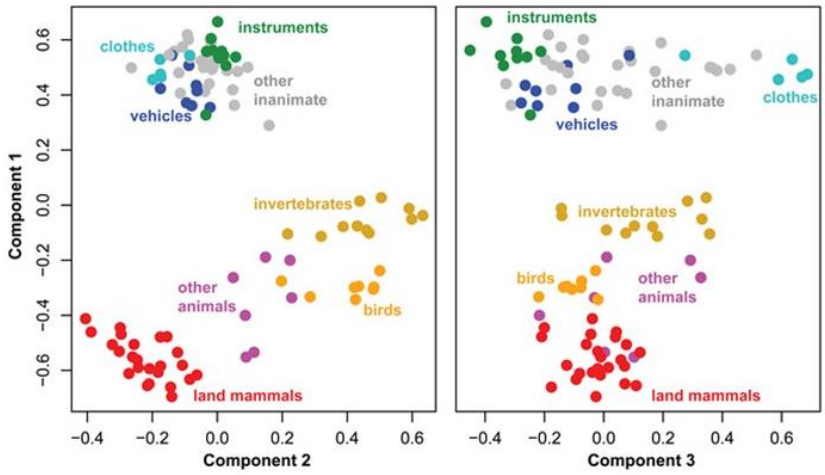

Figure 2: Coordinates of all stimuli on each target semantic dimension. Colours indicate category membership – land mammals (red), birds (orange), invertebrates (brown), other animals (magenta), instruments (green), clothes (light blue), vehicles (dark blue), and other inanimate objects (grey). Reprinted with permission from Cox and colleagues[28].

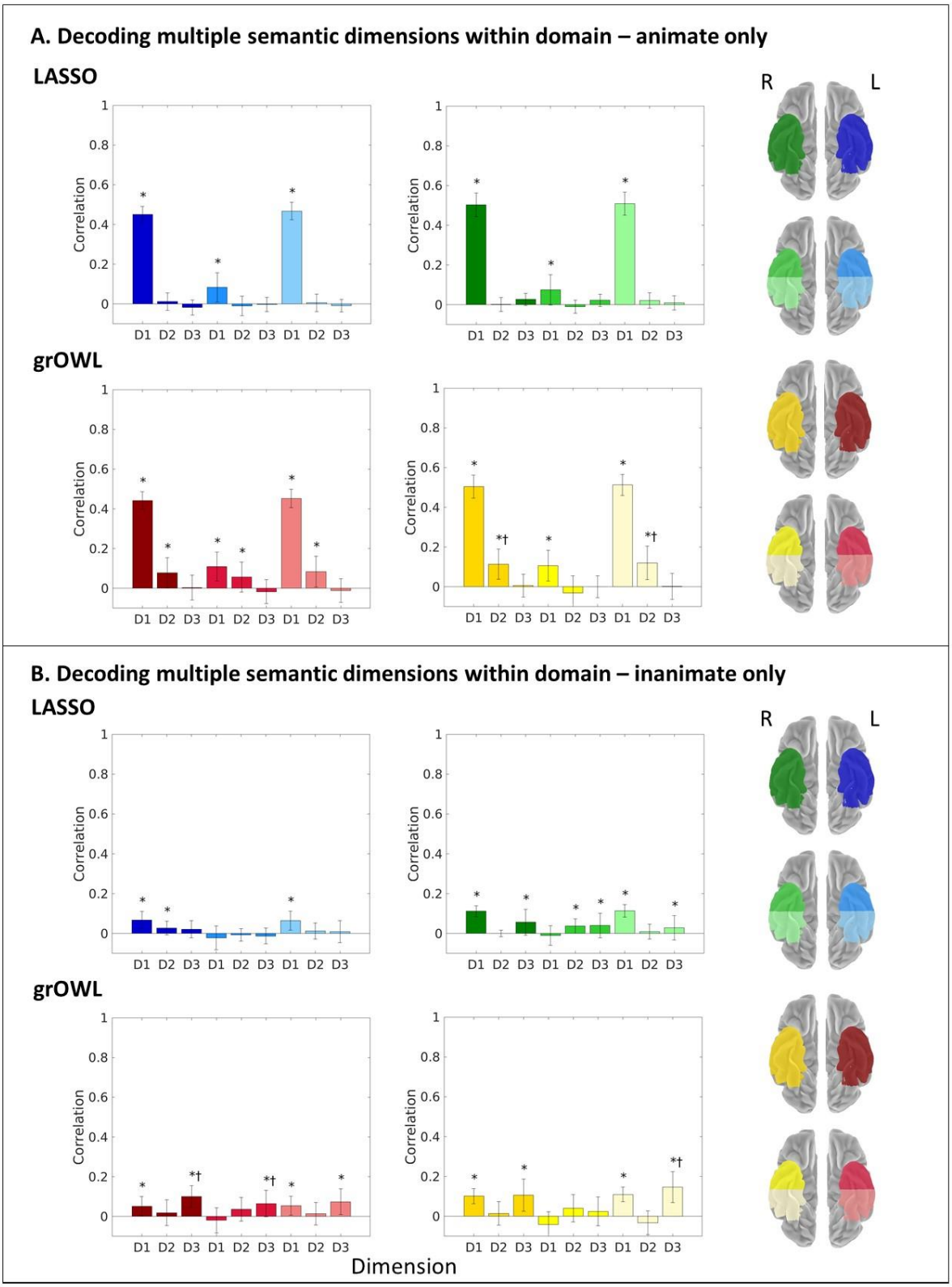

Figure 3: Hold-out correlations within-domain in regions of interest. Mean and 95% confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Surface plots (right) indicate which bar corresponds to which ROI or half-ROI. Models are trained using LASSO regularisation on the left ROI and half-ROIs (upper left, shades of blue), using LASSO

regularisation on the right ROI and half-ROIs (upper right, shades of green), using grOWL regularisation on the left ROI and half-ROIs (upper left, shades of red), or using grOWL regularisation on the right ROI and half-ROIs (upper right, shades of yellow). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension in the same ROI using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (A) Hold-out correlations for animate stimuli only; (B) Hold-out correlations for inanimate stimuli only.

2.2. Characterising semantic representations at the whole-brain level

All contemporary theories of sematic cognition implicate some regions outside the vATL – modality-specific regions[2,3,6–8], posterior temporal cortex[7,49,50], the inferior parietal lobe including the angular gyrus[1,2], or even the whole cortex[5,14,15]. Therefore, we next aimed to investigate where else in the brain semantic information could be detected and to characterise the semantic representations there. For these analyses, all models were trained on contrast beta values (one per stimulus, averaged over repeated presentations of the same stimulus) from the whole brain. Models were assessed using ten-fold cross-validation.

2.2.1. Is semantic structure present at the whole-brain level?

2.2.1.1. Animacy

Figure 4A shows the results from decoding animacy with logistic regression with LASSO or SOSLASSO regularisation. As in the ROI analysis, classification accuracy was very high (hold-out accuracy ≅ 0.95). This result confirmed the presence of information about animacy.

To assess which regions drove this result we visualised classifier coefficients on the cortical surface (see Methods) and compared these to univariate results (animate > inanimate and inanimate > animate; Figure 5A). Both univariate and multivariate analyses highlighted posterior temporal and occipitotemporal regions – the univariate map included some parietal regions that were absent from the multivariate maps and the multivariate maps encompassed the ventral posterior temporal lobe more comprehensively than the univariate map did.

371

372 2.2.1.2. Graded, multidimensional semantic structure

373

374 As shown in Figure 4B, decoding of dimension 1 was very successful ($r \cong 0.8$). However,
375 evidence for *multidimensional* structure depended on the regularisation penalty - models
376 with grOWL produced significant correlations on dimensions 2 and 3 whereas models with
377 LASSO did not. The results from RSL with grOWL provided evidence for multidimensional
378 semantic structure at the whole-brain level.

379 Figure 6 shows the hold-out correlations calculated for animate and inanimate
380 stimuli separately. Although only RSL with grOWL revealed multidimensional structure when
381 all stimuli were considered together, both RSL with LASSO and RSL with grOWL produced
382 significant correlations on two dimensions within-domain. For animate stimuli, these were
383 dimensions 1 and 2 (which separates out the animate stimuli) and, for inanimate stimuli, the
384 dimensions were dimension 1 and 3 (which separates out the inanimate stimuli). These
385 findings confirmed the presence of graded structure at the whole-brain level.

386 Figure 7 shows the coefficients from RSL models. As in the classification analyses,
387 posterior temporal and occipitotemporal regions were selected. Other regions that have
388 been proposed to represent semantic information, such as the angular gyrus[1,2], were absent
389 from these maps. The vATL was also absent[6,9,41], despite the fact that our ROI analysis
390 detected semantic structure there.

391

392 2.2.2. What are semantic representations outside the vATL like?

393

394 2.2.2.1. Categories, features, or vector-spaces?

395

396 Figure 4B shows that only RSL with grOWL was successful at decoding more than one
397 dimension when all stimuli were considered together; Figure 6 shows that RSL with LASSO
398 could successfully decode two dimensions within-domain, but was outperformed by RSL
399 with grOWL when decoding dimension 2 for animate stimuli and dimension 3 for inanimate
400 stimuli (permutation $p < 0.05$). Figure 7 shows that, when visualising coefficients on the
401 cortical surface, models with grOWL yielded significant results for all three dimensions.
402 However, for models with LASSO, coefficients for dimensions 2 and 3 did not survive

403 permutation thresholding. This clear advantage of grOWL over LASSO indicates that the
404 representations in posterior temporal cortex are likely to be interdependent, vector-space
405 representations[15].
406
407 2.2.2.2. Homogeneous or heterogeneous?
408
409 The coefficient maps for the classification analysis (Figures 6B and 6C) feature lateral
410 occipitotemporal regions in which coefficients are consistently negative across participants
411 (meaning that, in most participants, a positive beta value is associated with an increased
412 probability that the stimulus is animate) and medial occipitotemporal regions in which
413 coefficients are consistently positive. The same pattern is seen in Figure 7B in the
414 coefficients for dimension 1 in the RSL analysis with grOWL (this correspondence would be
415 expected because dimension 1 distinguishes animate from inanimate stimuli; Figure 2). This
416 demonstrates that these representations are heterogeneous – some units increase in
417 activity when representing a concept; others decrease.
418
419 2.2.2.3. Consistent or inconsistent?
420
421 The coefficient maps for the classification analysis (Figures 6B and 6C) and for dimension 1
422 in the RSL analysis with grOWL regularisation (Figure 7B) feature regions in which
423 coefficients are inconsistent in direction across participants. In the coefficient maps for
424 dimensions 2 and 3, the regions of inconsistent coefficient directions were even larger.
425
426 2.2.2.4. Contiguous or dispersed?
427
428 In contrast to the ROI analysis, there was no difference in classification accuracy between
429 SOSLASSO and LASSO (Figure 4A) and the same regions were highlighted with both methods
430 (Figures 6B and 6C). This does not constitute evidence for contiguity within and across
431 participants, although note also that classification performance may have been at ceiling
432 and that this may account for the similarity between methods.
433
434

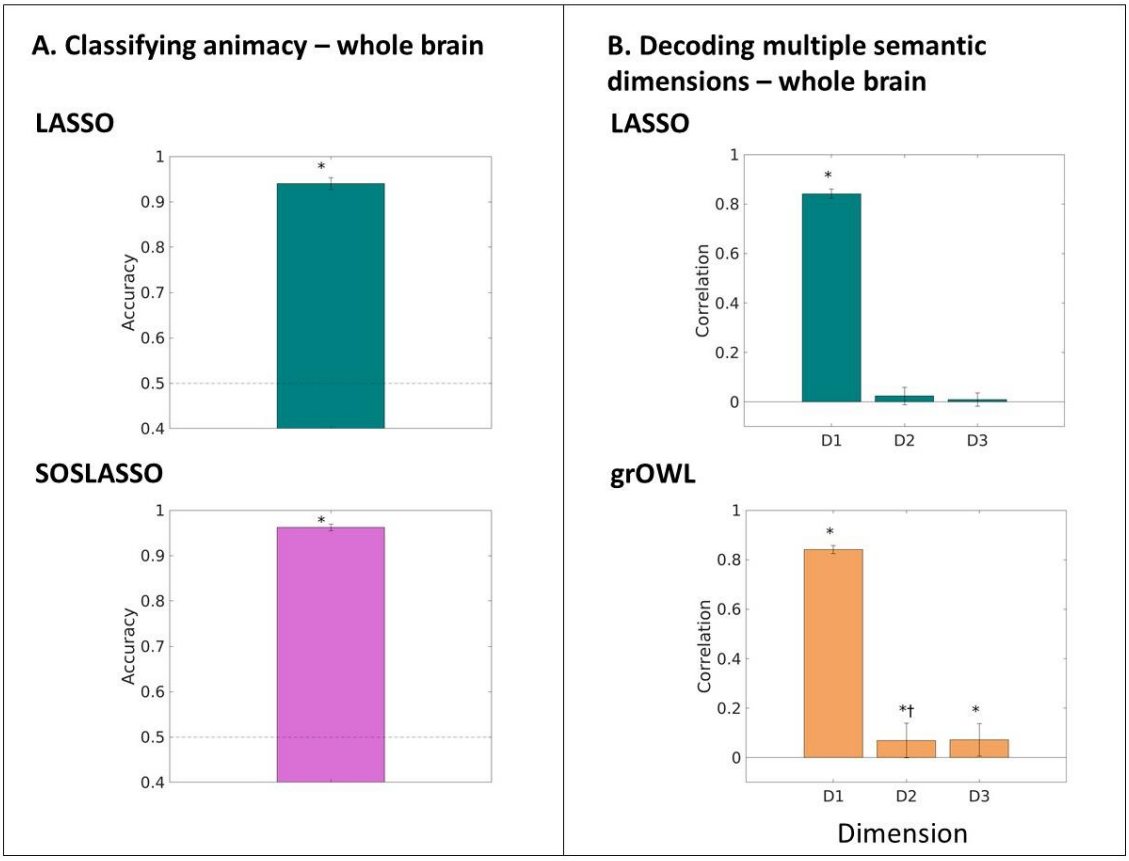

Figure 4: Decoding results at the whole-brain level. (A) Mean and 95% confidence interval of the hold-out accuracy for regularised logistic regression classifiers trained on contrast beta values to discriminate animate from inanimate stimuli. Classifiers are trained using LASSO regularisation (top, teal) or using SOSLASSO regularisation (bottom, magenta). Stars (*) indicate a significant difference between classifier accuracy and chance (as determined by permutation test, probabilities adjusted to control the false-discovery rate at α = 0.05). (B) Mean and 95% confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Models are trained using LASSO regularisation (top, teal) or using grOWL regularisation (bottom, orange). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at α = 0.05). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at α = 0.05).

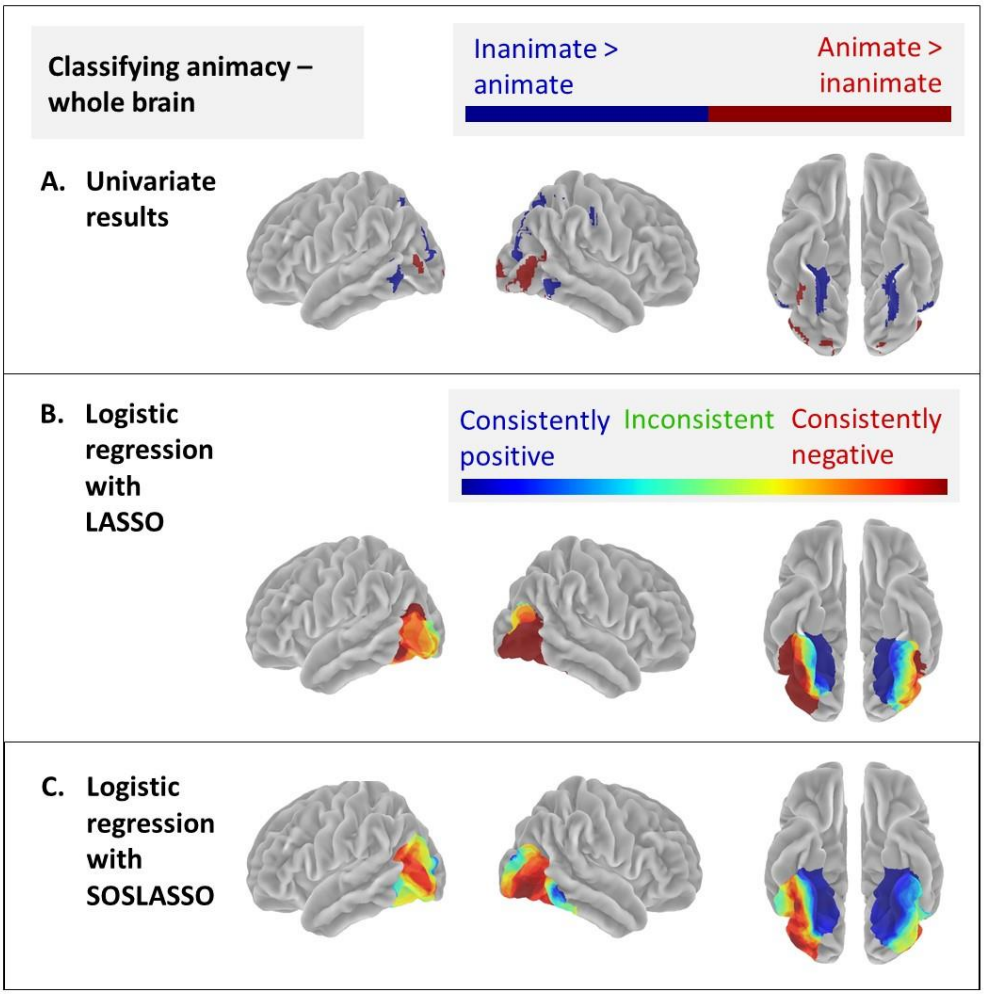

Figure 5: Coefficients for classification analyses. (A) Univariate effects (A > I, red; I > A, blue) are shown projected to the cortical surface and are cluster-corrected at $p < 0.05$ based on an uncorrected voxel threshold of $p < 0.001$. (B) Coefficients for logistic regression classifiers trained with LASSO regularisation on contrast beta values to discriminate animate from inanimate stimuli are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at $\alpha = 0.05$ (see Methods). Warm colours indicate that coefficients are mostly negative across participants (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that a positive beta value is associated with increased probability that the stimulus is animate); cool colours indicate that coefficients are mostly positive across participants (a positive beta value is associated with increased probability that the stimulus is inanimate); green shades indicate that the region was selected in multiple participants but that the sign of the coefficient was not consistent across participants. (C) Coefficients for logistic regression classifiers trained with SOSLASSO regularisation on contrast beta values to discriminate animate from inanimate stimuli are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at $\alpha = 0.05$ (see Methods). The colour scaling is the same as in (A).

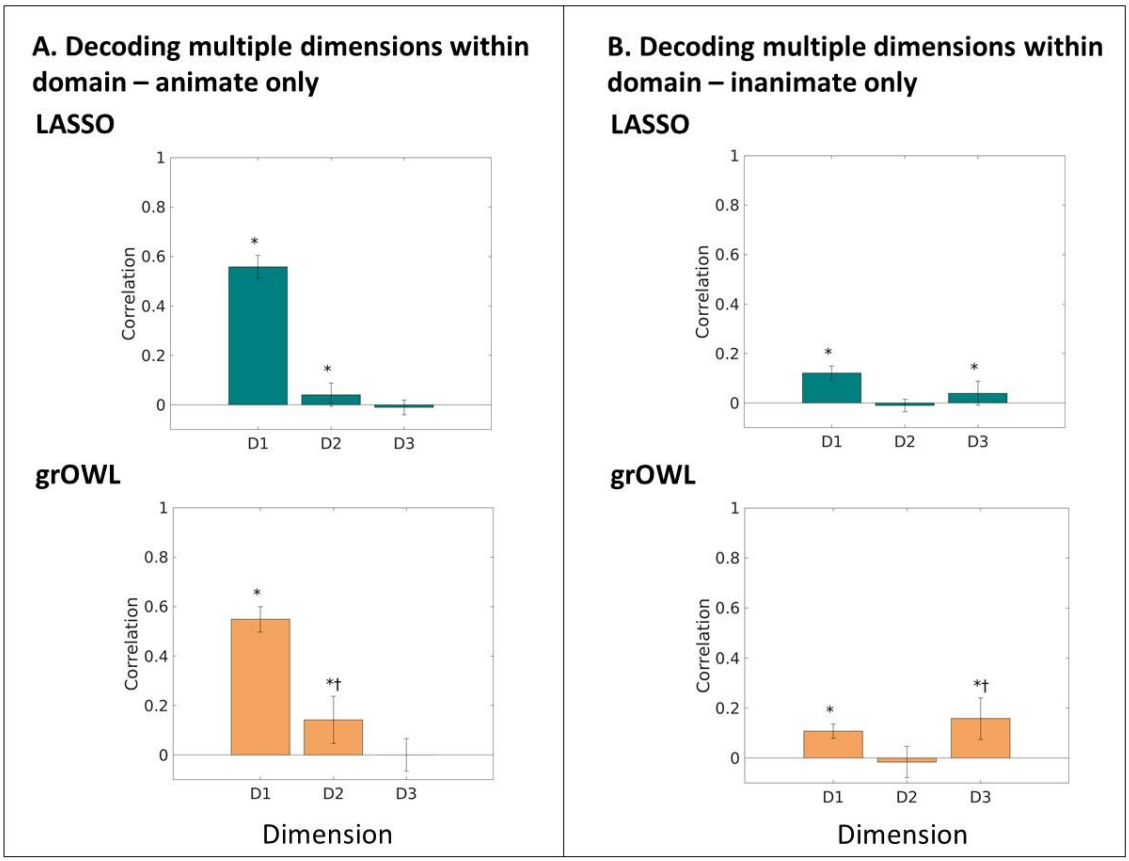

Figure 6: Hold-out correlations within-domain at the whole-brain level. Mean and 95% confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Models are trained using LASSO regularisation (top, teal) or using grOWL regularisation (bottom, orange). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at α = 0.05). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at α = 0.05). (A) Hold-out correlations for animate stimuli only; (B) Hold-out correlations for inanimate stimuli only.

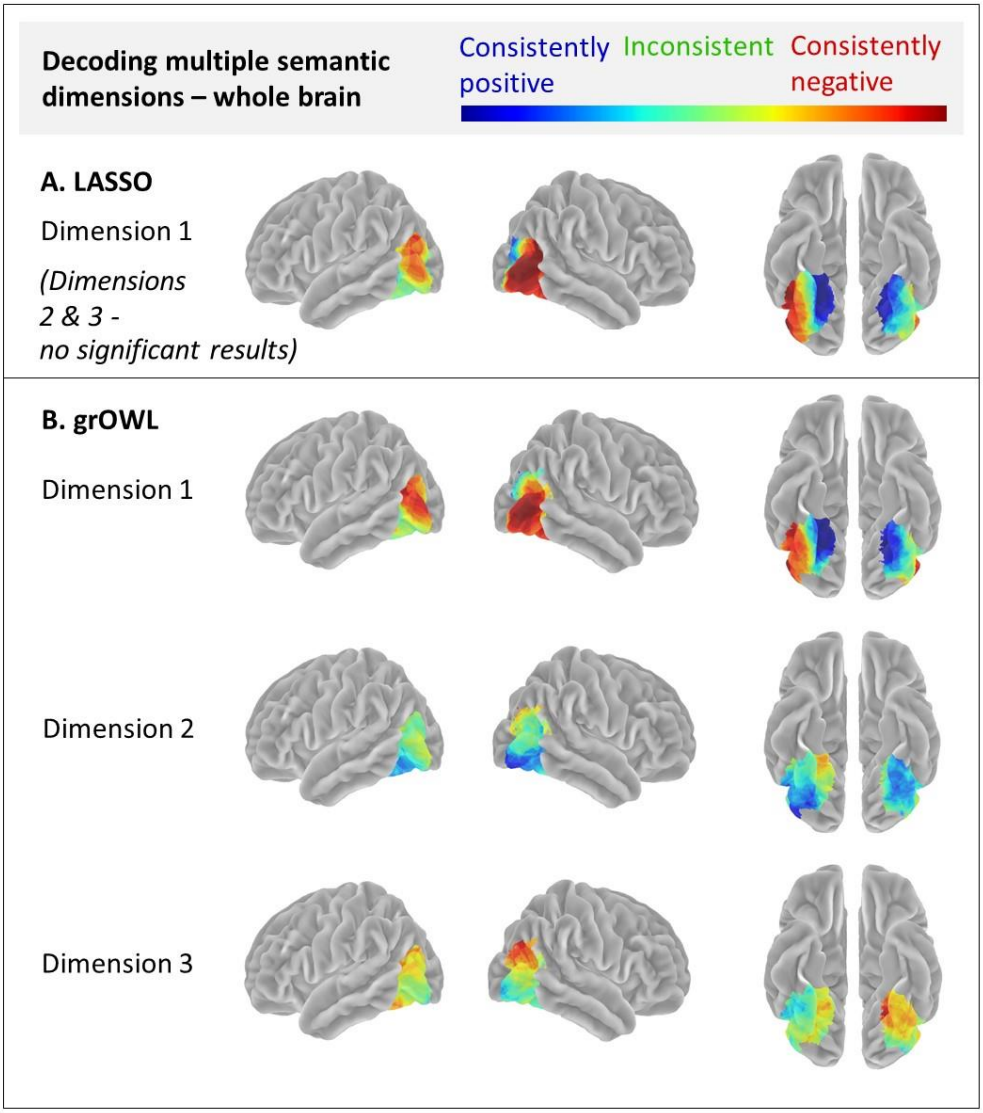

Figure 7: Coefficients for RSL analyses. (A) Coefficients for RSL models trained with LASSO regularisation on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at α = 0.05 (see Methods). Warm colours indicate that coefficients are mostly negative across participants; cool colours indicate that coefficients are mostly positive across participants; green shades indicate that the region was selected in multiple participants but that the sign of the coefficient wasnot consistent across participants. (B) Coefficients for RSL models trained with grOWL regularisation on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at α = 0.05 (see Methods). The colour scaling is the same as in (A).

504 **3. Discussion**

505

506 Equipped with a novel 7T-fMRI acquisition protocol capable of recording signal in ventral
507 temporal regions[26] and four innovative decoding methods[11,28–31], this study detected,
508 characterised, and thereby adjudicated theories about the brain's semantic representations.
509 We revealed graded, multidimensional semantic structure in the vATL, represented by an
510 interdependent, contiguous, dynamic[44] code. We also discovered interdependent,
511 heterogeneous, inconsistent[27] representations in posterior temporal and occipitotemporal
512 cortex.

513 Multiple theories of semantic cognition implicate the vATL, but those theories differ
514 in the neural code that they propose to exist in the vATL. Damasio and colleagues[3] hold that
515 the vATL contains convergence zones that simply coordinate representations elsewhere –
516 the vATL does not itself express semantic information. Conversely, we could decode
517 multiple semantic dimensions from vATL activity. Other theories argue that the vATL
518 functions as a hub that assembles and encodes complex semantic representations. The
519 finding that RSL with grOWL, which assumes that features are interdependent,
520 outperformed LASSO indicates that the vATL represents semantic information as an
521 uninterpretable vector-space code, aligning with predictions of the hub-and-spoke
522 model[6,9,41] rather than theories positing independent representational elements[1,2,40]. Also in
523 accordance with predictions of the hub-and-spoke model, poorer decoding performance
524 was observed in the anterior than the posterior half of our vATL ROI, suggesting the
525 existence of a dynamic code that is most clearly observed with time-resolved imaging
526 methods.

527 Our ROI results converge with results from neuropsychology[32–35], TMS[23,36–39],
528 intracranial electrophysiology[28,44,59–64], and computational modelling[9,42,43], yet diverge from
529 other multivariate studies using fMRI, which rarely highlight the vATL[10–18]. Two factors are
530 likely to account for this difference. The first is signal inhomogeneity caused by the
531 proximity of the vATL to the air-filled sinuses. Employing a novel multi-echo, multiband 7T-
532 fMRI acquisition sequence[26], in contrast to the standard acquisition used by others, we were
533 able to decode both animacy and graded, multidimensional semantic similarity structure
534 even from the most anterior region of the vATL. Collecting data with a standard sequence
535 (to enable a direct comparison) was beyond the scope of this study; previous work has

536 directly tested the link between acquisition sequence and correlation-based decoding
537 accuracy[26] and others should extend this test to decoding methods that depend on machine
538 learning (including logistic regression and RSL). The second factor is choice of method. This
539 study illustrated clear differences between decoding methods – SOSLASSO produced higher
540 classification accuracy than LASSO and grOWL decoded multiple semantic dimensions more
541 reliably than LASSO. The regularisation penalties that we used make assumptions (sparsity
542 and/or anatomical clustering and/or interdependence) that are explicitly based on theories
543 about how neural systems represent information[11,28]. However, note that, in comparison to
544 common approaches such as encoding[5,13,14] and RSA[16–18], the assumptions they make are
545 few and gentle. Our results suggest that adopting mild, neurally-inspired assumptions may
546 produce better coherence between fMRI results and results from other methods. However,
547 as argued previously[27], the most appropriate method for future studies is the one that aligns
548 with those individual researchers' explicit hypotheses. In summary, aided by two technical
549 advances, we have been able to reconcile the discrepancy between multivariate imaging
550 and other work.

551 Our whole-brain analysis discovered graded, multidimensional semantic similarity
552 structure in posterior temporal and occipitotemporal cortex. These regions have been
553 implicated by multiple theories – some propose that they encode visual properties including
554 colour, form, and motion[7,8] that may be particularly relevant for representing tools[49]; others
555 even propose that they encode mental states[5]. Our results enable us to distinguish
556 competing theories in two ways. First, as in the ROI analysis, we can adjudicate theories
557 based on the *nature* of the code that they propose. Detecting codes of varying natures is an
558 advantage of our multivariate approach[27] – note that the regions of inconsistent code
559 direction (green in Figures 5B and 5C) were missing from the univariate analysis (Figure 5A).
560 The interdependent, heterogeneous, inconsistent representations that we discovered
561 match the *kind* of code predicted by the hub-and-spoke model[6,9,41] (see also[15]). Note,
562 however, that the hub-and-spoke model does not predict what informational content we
563 should expect in the spokes (a set of regions which may include the posterior temporal and
564 occipitotemporal cortex). Preliminary evidence from computational modelling suggests that
565 spoke regions do come to represent overall semantic similarity structure (rather than
566 similarity structure within a particular modality) during semantic tasks[42,50], but it may only
567 be the spoke that is processing the current sensory input that does so[65]. The results of this

568 study are consistent with these exploratory findings – we presented visual stimuli and found
569 similarity structure in visual association cortex.
570 A second method of distinguishing between theories is considering which regions
571 were *not* selected in our analysis. The angular gyrus, which has been proposed as an
572 additional semantic hub[1,2], was absent, which means that our results do not readily support
573 a multi-hub perspective. The differences between this study and previous work[1,66,67],
574 however, remain to be explored in detail. One possible explanation is that the angular gyrus
575 is not a semantic store, but instead buffers multimodal and/or temporally-extended
576 information, including semantic information, during cognition[54,68]. Alternatively or
577 additionally, the angular gyrus may contribute to semantic control, possibly by representing
578 a semantic space that is "warped" to emphasise task-relevant dimensions[69]. To test this,
579 future studies should employ paradigms that make greater control requirements than our
580 naming paradigm does.
581 The vATL was also absent from our whole-brain results, despite the fact that our ROI
582 analysis detected semantic structure there. This discrepancy may arise because the
583 permutation distribution used for statistical thresholding of these maps is nonuniform –
584 voxels in the vATL are selected far more frequently than voxels elsewhere (Supplementary
585 Figures S1 and S2). This methodological quandary remains unresolved, but preliminary
586 simulation results suggest that overselection occurs when the correlated noise structure
587 present in fMRI breaks down under conditions of magnetic inhomogeneity, increasing the
588 dimensionality of the data and, consequently, the probability that data will correlate with a
589 permuted target (Frisby, unpublished analyses). Future work should both probe this
590 problem further and evaluate alternative methods of statistical significance testing for
591 decoding.
592
593 **5. Conclusion**
594
595 Previous theoretical work described the kinds of semantic representations that might
596 possibly exist[27]. This study represents the first endeavour to adjudicate those possibilities
597 empirically. We revealed graded, multidimensional semantic structure in the vATL,
598 represented by an interdependent, contiguous, dynamic[44] code. We also discovered
599 interdependent, heterogeneous, inconsistent[27] representations in posterior temporal and

occipitotemporal cortex. These results align better with the hub-and-spoke model of semantic cognition than with competing theories. Our comparative multivariate approach is an effective method that has the potential to reveal how the brain represents, not only semantic knowledge, but any kind of information.

6. Online methods

6.1. Participants

32 participants were recruited to the study. Of these, five were unable to complete the study either for medical reasons (for example, an incidental finding) or because of technical problems during data acquisition which meant that the scanner needed to be rebooted. This left 27 participants (age range 19-50, mean age 27.96 years, 17 female, 10 male) who were all native speakers of British English, were right-handed, had normal or corrected-to-normal vision, and had no neurological or sensory disorders. All participants gave written informed consent and the research was approved by a local National Health Service (NHS) ethics committee.

6.2. Stimuli and task

Stimuli were the same 100 line drawings used in previous ECoG work[28,44,59] – 50 animals and 50 inanimate items including vehicles, clothes, musical instruments and a range of other objects[70] (https://github.com/slfrisby/7TConvergent/tree/main/stimuli/). There were no significant differences between the categories with respect to age of acquisition, visual complexity, familiarity, word frequency, name agreement, and non-semantic visual structure[44,71]. PsychoPy 2022.2.5[72] was used to display stimuli and to record speech. Stimuli for both tasks were rear-projected onto a screen in the MRI scanner, and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes.

The study had a fast event-related design with 4 runs per participant, collected in a single session. Each run began 16 s after the start of the MR sequence and consisted of 100 trials (each stimulus appeared once per run). Each trial consisted of a stimulus presented for 4 s followed by a fixation cross presented for 4 s. 4 sequences of animate and inanimate stimuli were created and counterbalanced across participants. The order of trials was optimised for the contrast between animate and inanimate stimuli using optseq2

632 (https://surfer.nmr.mgh.harvard.edu/fswiki/optseq2). For each run and each participant, a
633 random order of animate stimuli and a random order of inanimate stimuli were created; the
634 stimuli were then interspersed according to the sequence generated with optseq2.
635 Participants were instructed to name each picture out loud as quickly and accurately
636 as possible while keeping head motion to a minimum. To ensure that participants
637 understood the instructions, they practised the task before entering the scanner while the
638 experimenter provided feedback about their head motion. Practice continued until the
639 experimenter was satisfied that head motion was minimised.
640 Overt speech was recorded for each trial and responses were scored for accuracy
641 with the study's aim of decoding semantic similarity structure in mind. If the participant did
642 not respond within the 4 seconds during which the stimulus was on-screen, the trial was
643 marked as incorrect. If the participant responded incorrectly but named an object that the
644 rater judged to be visually and semantically similar to the stimulus and at a similar level of
645 specificity (e.g. "wasp" for bee), the trial was marked as correct; if the object that they
646 named was semantically dissimilar or was named at a less specific level (e.g. "insect" for
647 bee), the trial was marked as incorrect. It was important that neural responses did not
648 reflect a "blend" of the semantics of multiple stimuli so, if the participant gave two
649 synonyms (e.g. "gun, pistol") the trial was marked as correct; if they named two different
650 objects ("kettle, iron") the trial was marked as incorrect even if one of the names was
651 correct. Trials containing fillers such as "um" and "err" in addition to a correct response
652 were marked as correct because the participants were neurologically healthy and there was
653 no reason to believe that these utterances reflected anything other than normal speech
654 processes. The mean and standard deviation of the accuracy scores were calculated.
655
656 6.3. Image acquisition
657
658 All images were acquired on a whole-body 7T Siemens MAGNETOM Terra MRI scanner
659 (Siemens Healthcare, Erlangen, Germany) with an 8Tx32Rx head coil (Nova Medical, USA).
660 An MP2RAGE anatomical scan was acquired with the following parameters: 224 sagittal
661 slices (interleaved acquisition), FOV 240 x 225.12 x 240 mm, voxel size 0.75 x 0.75 x 0.75
662 mm, TR 4300 ms, inversion times 840 and 2370 ms, TE 1.99 ms, flip angles 5° and 6°,
663 GRAPPA acceleration factor 3. Next, manual B_0-shimming was performed over the volume

664 to be used for echo-planar imaging (EPI). The aim was to reduce the water linewidth,
665 defined as full width of the spectrum at half height, to below 40 Hz. However, for some
666 participants the adjustments proved time-consuming and so adjustment time was capped at
667 30 minutes from the start of scanning after which the best shim parameters were adopted.
668 The actual average water linewidth was 38.80 Hz (standard deviation = 5.93 Hz).
669 All four functional runs of EPI used the multi-echo multiband sequence optimised by
670 Frisby and colleagues[26] with the following parameters: 48 axial slices (interleaved
671 acquisition), FOV 210 x 210 x 210 mm to cover the whole brain in most participants (visual
672 inspection ensured that the vATL was included in the FOV for all participants and the FOV
673 was tilted up at the nose to avoid ghosting of the eyes into the vATL), TR = 1510 ms, 3
674 echoes, TE_1 = 11.80 ms, TE_2 = 27.05 ms, TE_3 = 42.30 ms, multiband factor 2, GRAPPA
675 acceleration factor 3, phase partial Fourier 7/8, voxel size 2.5 mm isotropic (no gap), flip
676 angle 63°, and A-P phase encoding direction. Immediately *before* each run, 5 additional
677 volumes were acquired with the phase encoding direction changed to P-A. These volumes
678 were to be used for distortion correction during preprocessing.
679
680 ### 6.4. Data analysis
681
682 All analysis code is available at: https://github.com/slfrisby/7TConvergent/.
683
684 #### 6.4.1. Preprocessing
685
686 For ease of data sharing we converted all DICOMs to BIDS format[73] using *heudiconv*
687 (v1.1.1)[74].
688 Standard reproducible preprocessing pipelines designed for 3T-fMRI, such as
689 *fMRIprep*[75–77], perform poorly on 7T-fMRI EPI data; therefore, the analysis pipeline was split
690 into stages using different software packages. The MP2RAGE T1w (combined) image first
691 had its background noise removed using O'Brien regularisation[78] and was then submitted to
692 CAT12 for segmentation[79] (in SPM12; https://www.fil.ion.ucl.ac.uk/spm/). The bias-
693 corrected and global-intensity-corrected T1w (combined) image produced was provided to
694 as input to *fMRIPrep* 21.0.1. The T1w (combined) image was corrected for intensity non-
695 uniformity with *N4BiasFieldCorrection*[80] and skull-stripped with a Nipype implementation of

696 the *antsBrainExtraction.sh* workflow (ANTs). Brain tissue segmentation of cerebrospinal
697 fluid (CSF), white matter (WM), and grey matter (GM) was performed on the brain-extracted
698 T1w (combined) image using *fast* (FSL 6.0.5.1)[81]. Volume-based spatial normalisation to
699 standard space (MNI152NLin2009cAsym) was performed through nonlinear registration
700 with *antsRegistration.sh* (ANTs 2.3.3, https://github.com/ANTsX/ANTs)[82,83].
701 Functional preprocessing was performed using in-house code, composed of
702 functions from AFNI (v.18.3.03; https://afni.nimh.nih.gov/)[84,85], FSL (v.5.0.10;
703 https://fsl.fmrib.ox.ac.uk/fsl/fslwiki)[86–89], tedana (v.24.0.0;
704 https://tedana.readthedocs.io/en/stable/index.html)[90–93], and ANTS (v. 2.2.0;
705 https://github.com/ANTsX/ANTs)[82,83]. EPIs were slice-time corrected to the middle slice
706 using *3dTshift* (AFNI). Next, EPIs from the first echo were motion corrected with *3dvolreg*
707 and *3dAllineate* using the first volume of the first run acquired as a reference. The same
708 transforms were applied to the EPIs from the second and third echoes. EPIs were then skull-
709 stripped using *BET* (FSL)[88] to create a participant-specific brain mask. *tedana* was used to
710 combine data from all echoes optimally based on T2* weighting[94] and to denoise the data
711 using ICA. *tedana* conducts denoising by decomposing data using PCA and ICA, classifying
712 components according to whether the signal scales linearly with TE (as BOLD signal does),
713 and reconstructing the data using only BOLD-like components. The brain mask created with
714 *BET* was used as the mask for this stage.
715 Next, unwarping was conducted using *topup* and *applytopup* (FSL)[95]. Field
716 displacement maps were calculated using 10 volumes (5 with A-P phase encoding direction,
717 extracted from the start of each functional run, and 5 with P-A phase encoding direction,
718 collected separately before each run) and the resulting correction was applied to all images.
719 Finally, the mean EPI for each run was coregistered to the skull-stripped native
720 structural image using a rigid-body registration in *AntsRegistrationSyN.sh* (ANTS). This
721 produced transforms from native EPI space to native T1 space. For all decoding analyses,
722 images remained in native space and were not smoothed. However, the transforms from
723 native EPI to native T1 (for the first run) and the transforms from native T1 to standard
724 space (MNI152NLin2009cAsym), calculated with *antsApplyTransforms* (ANTS), were used to
725 calculate the MNI-space coordinates that each voxel *would* have *were* images to be
726 transformed into MNI space. (These coordinates were used for visualisation and were
727 provided to classifiers trained with SOSLASSO regularisation, which take anatomical

information as input.) The transform from native EPI to native T1 space was also used to transform the participant's grey-matter mask into native EPI space. For the univariate analysis, images were transformed into standard space (MNI152NLin2009cAsym) with *antsApplyTransforms* (ANTS) and smoothed with a 6 mm FWHM Gaussian filter in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/).

6.4.2. 1st-level (within-participant) GLM

Data were processed using the general linear model (GLM) approach implemented in SPM12 in MATLAB r2023b. For decoding, each stimulus was modelled as an individual boxcar function, 1.5 seconds in length because the stimulus in our stimulus set with the longest mean naming time had a mean naming time of 1323 ms[70] (periods between stimuli were modelled implicitly). These boxcar functions were subsequently convolved with SPM's difference of gammas haemodynamic response function. The six motion parameters extracted during preprocessing were used as regressors of no interest. Data from all four runs were concatenated, so an additional constant was included for each run. The micro-time resolution was set as the number of slices (n = 48), the micro-time onset was set as the reference slice for slice-time correction (n = 24), and the high-pass filter cutoff was 128 seconds. Each participant's grey matter mask in native EPI space was used as a mask for the analysis. The parameter estimation method was restricted maximum likelihood estimation (ReML) and serial correlations were accounted for using an autoregressive AR(1) model during estimation. This produced one beta image per stimulus in native space. For univariate analysis, the same model was run on the normalised, smoothed data with identical parameters except that a grey matter mask in standard space (MNI152NLin2009cAsym) was used and that stimuli named incorrectly were modelled as additional regressors of no interest. The univariate contrasts of interest were greater activation for animate stimuli named correctly than inanimate stimuli named correctly (A > I) and greater activation for inanimate stimuli named correctly than animate stimuli named correctly (I > A). This produced contrast images in MNI space.

6.4.3. Univariate analysis – 2nd-level (across-participant) GLM

760 Group-level t-tests of each contrast of interest (A > I, I > A) were conducted in SPM12. The
761 group t-maps were assessed for significance by using a voxel-height threshold of $p < 0.001$
762 to define clusters and then a cluster-defining family-wise-error-corrected threshold of $p <$
763 0.05 for statistical inference. Group-level results were then projected to the pial surface
764 (fsaverage template[96,97]) using *-volume-to-surface-mapping* with trilinear interpolation in
765 Connectome Workbench 1.5.0 and the thresholded maps were plotted on the pial surface
766 using *plot_surf_stat_map* in *nilearn* implemented in Python 3.9.
767
768 6.4.4. Multivariate decoding
769
770 To prepare data for decoding, beta images from all 4 presentations of each stimulus
771 were averaged across runs. If a participant had named a stimulus incorrectly on one to three
772 occasions, the beta image(s) corresponding to the incorrect trial(s) were not included in the
773 averaging. If the participant named the stimulus incorrectly on all four trials, the values for
774 each voxel were interpolated with the median value for that voxel over all other stimuli.
775 We constructed our vATL ROI based on electrode coverage in a previous study[59] (all
776 electrode coordinates are available at https://github.com/slfrisby/7TConvergent/).
777 Electrode coordinates in the temporal lobe were converted to a single NIFTI volume at
778 native resolution (2.5 mm isotropic) in MNI space using SPM12 and smoothed at 10mm
779 FWHM. In order to exclude regions that lay outside the temporal lobe, a mask was created
780 by taking the Harvard-Oxford cortical and subcortical atlases, removing the temporal lobe
781 from the atlas, and adding the cerebellum from the AAL 2 atlas[98] (since the Harvard-Oxford
782 atlas does not include a cerebellum). Only voxels that lay outside this atlas mask were
783 included in the ROI. After masking the ROI was smoothed again at 4mm FWHM because, on
784 inspection, this created a smooth shape with space around the temporal lobes – since the
785 ROIs were to be backprojected into native space, it was important to ensure that individual
786 participants' temporal lobes would be fully encompassed. Since electrode coverage was far
787 more extensive in the left than the right hemisphere, a mirror-image of the left ROI was
788 used as the right ROI. To create half-ROIs to enable tests for dynamism, we bisected the
789 ROIs between their anteriormost and posteriormost points (y = -20). The final ROIs in
790 standard space (MNI152NLin2009cAsym) are shown projected to the surface in Figures 1
791 and 3 and are available at https://github.com/slfrisby/7TConvergent/tree/main/ROI/.

The ROIs were backprojected into each participant's native space using the transforms from native EPI to native T1 and the transforms from native T1 to standard space generated with *antsRegistration.sh*. For the ROI decoding analysis, only beta values from voxels that overlapped with both the ROI and the participant's native-space grey matter mask were provided as input to decoding models. (For the whole-brain analysis, only voxels within the participant's native-space grey matter mask were provided as input).

We tested for semantic structure using two decoding methods.

6.4.4.1. Regularised logistic regression

We used regularised logistic regression to decode animacy. Regularised logistic regression classifiers avoid overfitting by finding a set of coefficients β (one per feature, in this case per voxel) that minimise both prediction error and a regularisation penalty:

$$\min_{\beta}((1-\lambda)f(\beta)+\lambda h(\beta))$$

where $f(\beta)$ is the prediction error, $h(\beta)$ is the regularisation penalty, and λ is a learnt hyperparameter that controls how the optimisation should weigh the regularisation penalty versus the prediction error.

One set of classifiers was trained using LASSO (L1) regularisation[55], for which the regularisation penalty is:

$$h(\beta)=\sum_{j=1}^{m}|\beta_j|$$

The LASSO penalty increases as the sum of the absolute values of the coefficients increases and is therefore at its lowest when most coefficients are zero and the voxels that receive the few nonzero coefficients are uncorrelated in their activity. Classifiers with LASSO therefore encapsulate the assumption that the neural code is sparse (only a few voxels encode the target structure) and uncorrelated.

Classifiers with LASSO were fitted using *glmnet* in MATLAB r2023b[99], using 3 – 0.2 as the range of possible lambda values. Each classifier took, as input, beta values from multiple voxels from a single participant.

A second set of classifiers was trained using SOSLASSO regularisation[11,30,31]. Before model training, voxels are assigned to "sets" – ROI-like spheres defined in MNI space. In this

study, the sets were 9 mm in radius and overlapped by 9mm with adjacent spheres. All decoding analyses were conducted in native space, but each voxel was assigned to sets based on the location that it *would* have *were* the images to be transformed into MNI space (see Preprocessing). The sets are defined in a common space, so voxels from multiple participants are assigned to the same anatomical set.

Classifiers are then fit to beta values from multiple voxels from *all* participants. The SOSLASSO regularisation penalty is calculated for each set individually and summed over all sets S:

$$h(S,\beta) = \sum_{S}\left((1-\alpha)\sum_{j=1}^{m}|\beta_j| + \alpha\sqrt{\sum_{j=1}^{m}\beta^2} \right)$$

where α is a learnt hyperparameter that scales the two components of the SOSLASSO penalty – the LASSO penalty $\sum_{j=1}^{m}|\beta_j|$, which increases as the sum of the coefficients increases, and the grouping penalty $\sqrt{\sum_{j=1}^{m}\beta^2}$, which also increases as the sum of the coefficients increases but increases more steeply when those voxels are in different sets than when they are in the same set. This means that classifiers with SOSLASSO assume that the code is sparse, but also found in similar locations within and across participants.

Classifiers with SOSLASSO were fitted using the WISC MVPA toolbox in MATLAB r2018b (https://github.com/crcox/WISC_MVPA/) with hyperband budget set to 25 (all other parameters were set as defaults, including 3 – 0.2 as the range of possible lambda values).

6.4.4.2. Representational similarity learning

Next, we used RSL to decode graded, multidimensional similarity structure. We first constructed a representational similarity matrix (RSM) $\boldsymbol{S}$ that expressed the semantic similarity between each possible pair of stimuli (operationalised using feature verification norms[56]. We then applied singular value decomposition to extract three components accounting for 89.5% of the variance in the whole RSM (81.1%, 4.4%, and 4.0%, respectively)[28]. Figure 2 shows the coordinates of all 100 stimuli on each dimension. Where n is the number of stimuli and r is the number of dimensions, SVD yields three products:

1. An $n \times r$ matrix ($\boldsymbol{U}$) of components, also known as singular vectors.
2. The transpose of $\boldsymbol{U}$ ($\boldsymbol{U^T}$).
3. An $r \times r$ matrix ($\boldsymbol{D}$) of singular values.

To summarise:

$$\boldsymbol{S} \cong \boldsymbol{UDU^T}$$

We weighted the components $\boldsymbol{U}$ by the square root of the singular values $\sqrt{\boldsymbol{D}}$ to create an $n \times r$ matrix $\boldsymbol{C}$ – three target dimensions for decoding.

$$\boldsymbol{C} = \boldsymbol{U}\sqrt{\boldsymbol{D}}$$

Next, RSL models were trained to predict the coordinates of each stimulus on each dimension. Each voxel receives three coefficients – one per dimension – and so the set of coefficients to be learnt is an $m \times r$ matrix $\boldsymbol{\beta}$. Where $\boldsymbol{X}$ is the matrix of data (of size n stimuli x m voxels) and α is a learnt hyperparameter, RSL aims to find the set of coefficients that minimises:

$$\min_{\beta}((1-\lambda)\|\boldsymbol{C} - \boldsymbol{X\beta}\|_F + \lambda H(\boldsymbol{\beta}))$$

One set of models used LASSO regularisation and were trained using *glmnet* in MATLAB r2023b, using 6 – 0 as the range of possible lambda values. Each model took, as input, beta values from multiple voxels from a single participant.

A second set of models used grOWL regularisation. grOWL works by creating a vector w of non-negative, non-increasing values (i.e. each successive value of w is smaller than or equal to the previous value). w is of length m (the number of voxels) and a correspondence is learnt between elements of w and rows of $\boldsymbol{\beta}$ such that the row of $\boldsymbol{\beta}$ containing the largest coefficients (defined by the Euclidean norm or 2-norm of those coefficients) receives the largest value in w, the row containing the next largest coefficients receives the next largest value, and so on. Note that, because successive elements of w can be equal, rows of $\boldsymbol{\beta}$ containing very similar coefficients can receive the same value in w.

The total grOWL regularisation penalty is calculated by taking the Euclidean norm (also known as the 2-norm) of each row of $\boldsymbol{\beta}$, multiplying that with its corresponding value in w, and summing the results. Where $\boldsymbol{\beta}_{[i]\cdot}$ is the row of $\boldsymbol{\beta}$ with the i-th largest 2-norm and w_i is the value in w corresponding to $\boldsymbol{\beta}_{[i]\cdot}$:

$$H(\beta) = \sum_{i=1}^{m} w_i \left\| \boldsymbol{\beta}_{[i]\cdot} \right\|_2$$

$\left\| \boldsymbol{\beta}_{[i]\cdot} \right\|_2$ increases as each coefficient increases. However, unlike LASSO which prefers voxels to be uncorrelated in their activity, grOWL assigns similar values in w to correlated voxels, meaning that those correlated voxels either all receive nonzero coefficients or all receive coefficients of zero. Consequently, grOWL assumes that the neural code is sparse but correlated – in other words, while LASSO assumes that only a few voxels will be important, grOWL assumes that only a few *groups of correlated voxels* will be important. Note also that $\left\| \boldsymbol{\beta}_{[i]\cdot} \right\|_2$ is smaller when coefficients in a row of $\boldsymbol{\beta}$ are similar in size than when one is large and others are small or zero. grOWL therefore makes the *spanning assumption* – that important voxels should predict multiple target dimensions at once, and therefore that the target dimensions are represented interdependently.

grOWL models were trained using the WISC MVPA toolbox in MATLAB r2018a with hyperband budget set to 25 (all other parameters were set as defaults, including 6 – 0 as the range of possible lambda values). Each model took, as input, beta values from multiple voxels from a single participant.

6.4.4.3. Cross-validation and model assessment

Classifier or model accuracy was assessed via ten-fold nested cross-validation. In each outer loop, ten out of 100 stimuli (five animate, five inanimate) were defined as the outer-loop hold-out set. A second ten stimuli (five animate, five inanimate) were defined as the inner-loop hold-out set. The remaining 80 stimuli were used as training data to search for the best hyperparameter(s) - λ, plus α for SOSLASSO and w for grOWL. Hyperparameter performance was assessed on the inner-loop hold-out set. The folds were then reassigned so that ten stimuli previously used for training became the inner-loop hold-out set. Once all combinations of inner-loop hold-out set and training data had been explored, the best-performing hyperparameter value(s) were used to train a final model that was assessed on the outer-loop hold-out set. The procedure was then repeated with the nine other possible hold-out sets. In the first instance, stimuli were assigned randomly to the ten hold-out folds,

but this order was then fixed so that models fitted with every kind of regularisation were assessed using the same fold configuration.

The mean classification accuracy (for regularised logistic regression classifiers) or correlation between true and predicted coordinates (for RSL models) was calculated over all hold-out folds. Statistical significance was assessed via permutation testing. This was done for two reasons. First, since we used a fast event-related design, the haemodynamic response overlapped between trials. This means that, if temporally adjacent stimuli were divided between training and test sets, the stimulus in the training set may contain information about the stimulus in the test set and so may produce classification hold-out accuracies above 0.5 or hold-out correlations above 0. Second, cross-validated correlation (as was calculated for RSL models) is known to produce predicted coordinates that sometimes correlate *negatively* with the target coordinates in a hold-out set – the mean correlation under the null hypothesis may be below zero[28,100]; therefore, the magnitude of the real correlation can be understood only with reference to the permutation distribution. In permutation testing, the stimulus labels were permuted before model training (as for hold-out folds, the permutation orders were generated at random but then fixed so that the permutations for each model were comparable). The procedure was repeated 100 times to produce 100 simulated accuracy or correlation values for each participant. A group-level empirical null distribution was simulated by randomly sampling one of the 100 values from each participant and computing a group average 10,000 times[101]. One-tailed p-values were calculated (where m is the number of values in the permutation distribution and b is the number of values in the permutation distribution larger than the true value[28,102]):

$$p = \frac{b+1}{m+1}$$

To test whether one model or ROI produced a higher accuracy or correlation than another, a group-level permutation distribution of differences was created using the differences between the values in both permutation distributions and using the group-level simulation procedure described above. A p-value was then calculated for the true difference.

Statistical significance for all tests was defined after adjusting p-values to control the false discovery rate at $\alpha = 0.05$[103].

Ten-fold nested cross-validation does not provide a clear picture of which voxels are informative – each of the ten final models selects its own set of coefficients[11]. Therefore, to visualise coefficients, one model was trained for each participant using an inner loop of ten stimuli to identify the best hyperparameter(s) but no outer loop (i.e. all 90 stimuli that were not in the current inner loop hold-out set were used for hyperparameter exploration; therefore, all 100 stimuli were used for training). 100 permutation models were also trained for each participant. Coefficients were converted to one NIFTI volume per participant at native resolution (2.5 mm isotropic) in MNI space using SPM12. Coefficients both for models trained on real data and for models trained on permuted data were projected to the pial surface (fsaverage template[96,97]) using *-volume-to-surface-mapping* with trilinear interpolation and smoothed on the surface at 6 mm FWHM using *–metric-smoothing* in Connectome Workbench 1.5.0. Two metrics were calculated – (1) the proportion of participants in which each vertex was selected and (2) the proportion of negative coefficients that each vertex received (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that a positive beta value is associated with increased probability that the stimulus is animate). The map of the proportion of negative coefficients was thresholded using the map of coefficient selection - vertices were included in the map only if they were selected significantly more frequently in the real distribution than they were in the permutation distribution, as assessed via binomial tests with p-values adjusted to control the false discovery rate at $\alpha = 0.05$[103]. Finally, the thresholded map was plotted on the pial surface using *plot_surf_stat_map* in *nilearn* implemented in Python 3.9.

Data and code availability: Data will be made publicly available upon peer review and acceptance. Code is publicly available at: https://github.com/slfrisby/7TConvergent/.

Author contributions: Saskia L. Frisby – Conceptualisation, Methodology, Investigation, Formal Analysis, Visualisation, Writing – Original Draft, Writing – Review and Editing, Funding Acquisition; Christopher R. Cox – Conceptualisation, Methodology, Software, Writing – Review and Editing; Ajay D. Halai – Conceptualisation, Writing – Review and Editing, Supervision, Funding Acquisition; Matthew A. Lambon Ralph – Conceptualisation, Writing – Review and Editing, Supervision, Funding Acquisition; Timothy T. Rogers – Conceptualisation, Writing – Review and Editing, Supervision, Funding Acquisition.

Funding: This work was supported by an EPS study visit grant (G118497) to S.L.F., by an MRC Career Development Award (MR/V031481/1) to A.D.H., by an MRC programme grant (MR/R023883/1) and intramural funding (MC_UU_00005/18) to M.A.L.R., and by an NSF grant (NCS-FO 219903 DRL) to T.T.R.

Competing interests: The authors declare no conflicts of interest.

Acknowledgements: We would like to thank the participants and the Wolfson Brain Imaging Centre radiographers.

References

1. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex N. Y. N 1991* **19**, 2767–2796 (2009).
2. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).
3. Damasio, A. R. The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* **1**, 123–132 (1989).
4. Damasio, H., Tranel, D., Grabowski, T., Adolphs, R. & Damasio, A. Neural systems behind word and concept retrieval. *Cognition* **92**, 179–229 (2004).
5. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
6. Lambon Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* **18**, 42–55 (2017).
7. Martin, A. The representation of object concepts in the brain. *Annu. Rev. Psychol.* **58**, 25–45 (2007).
8. Martin, A. GRAPES — Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychon. Bull. Rev.* **23**, 979–990 (2016).
9. Rogers, T. T. *et al.* Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol. Rev.* **111**, 205–235 (2004).
10. Anderson, A. J. *et al.* Multiple Regions of a Cortical Network Commonly Encode the Meaning of Words in Multiple Grammatical Positions of Read Sentences. *Cereb. Cortex* **29**, 2396–2411 (2019).
11. Cox, C. R. & Rogers, T. T. Finding distributed needles in neural haystacks. *J. Neurosci.* **41**, 1019–1032 (2021).
12. Dehghani, M. *et al.* Decoding the neural representation of story meanings across languages. *Hum. Brain Mapp.* **38**, 6096–6106 (2017).
13. Fernandino, L. *et al.* Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cereb. Cortex* **26**, 2018–2034 (2016).

14. Popham, S. F. *et al.* Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat. Neurosci.* **24**, 1628–1636 (2021).
15. Pereira, F. *et al.* Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* **9**, 1–13 (2018).
16. Devereux, B. J., Clarke, A. & Tyler, L. K. Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Sci. Rep.* **8**, 10636 (2018).
17. Devereux, B. J., Clarke, A., Marouchos, A. & Tyler, L. K. Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *J. Neurosci.* **33**, 18906–18916 (2013).
18. Connolly, A. C. *et al.* The representation of biological classes in the human brain. *J. Neurosci.* **32**, 2608–2618 (2012).
19. Devlin, J. T. *et al.* Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *NeuroImage* **11**, 589–600 (2000).
20. Halai, A. D., Welbourne, S. R., Embleton, K. & Parkes, L. M. A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Hum. Brain Mapp.* **35**, 4118–4128 (2014).
21. Halai, A. D., Parkes, L. M. & Welbourne, S. R. Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *NeuroImage* **122**, 214–221 (2015).
22. Halai, A. D., Henson, R. N., Finoia, P. & Correia, M. M. Comparing the effect of multi gradient echo and multi band fMRI during a semantic task. 2024.03.20.585909 Preprint at https://doi.org/10.1101/2024.03.20.585909 (2024).
23. Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M. & Lambon Ralph, M. A. 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**, 2728–2738 (2010).
24. Embleton, K. V., Haroon, H. A., Morris, D. M., Lambon Ralph, M. A. L. & Parker, G. J. M. Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Hum. Brain Mapp.* **31**, 1570–1587 (2010).
25. Visser, M., Jefferies, E. & Lambon Ralph, M. A. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J. Cogn. Neurosci.* **22**, 1083–1094 (2010).
26. Frisby, S. L. *et al.* Optimising 7T-fMRI for imaging regions of magnetic susceptibility. 2025.03.17.643748 Preprint at https://doi.org/10.1101/2025.03.17.643748 (2025).
27. Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A. & Rogers, T. T. Decoding semantic representations in mind and brain. *Trends Cogn. Sci.* **27**, 258–281 (2023).
28. Cox, C. R. *et al.* Representational similarity learning reveals a graded multidimensional semantic space in the human anterior temporal cortex. *Imaging Neurosci.* **2**, 1–22 (2024).
29. Oswal, U., Cox, C. R., Lambon Ralph, M. A., Rogers, T. T. & Nowak, R. D. Representational Similarity Learning with Application to Brain Networks. in *Proceedings of The 33rd International Conference on Machine Learning* 1041–1049 (PMLR, 2016).
30. Rao, N., Nowak, R. D., Cox, C. & Rogers, T. Classification with the sparse group lasso. *IEEE Trans. Signal Process.* **64**, 448–463 (2016).
31. Rao, N., Cox, C., Nowak, R. D. & Rogers, T. T. Sparse overlapping sets lasso for multitask learning and its application to fMRI analysis. in *Advances in Neural Information Processing Systems* vol. 26 (2013).
32. Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P. & Hodges, J. R. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* **38**, 1207–1215 (2000).
33. Hodges, J. R. & Patterson, K. Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurol.* **6**, 1004–1014 (2007).
34. Snowden, J., Goulding, P. J. & Neary, D. Semantic dementia: A form of circumscribed cerebral atrophy. *Behav. Neurol.* **2**, 124043 (1989).

35. Warrington, E. K. The selective impairment of semantic memory. *Q. J. Exp. Psychol.* **27**, 635–657 (1975).
36. Lambon Ralph, M. A., Pobric, G. & Jefferies, E. Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cereb. Cortex* **19**, 832–838 (2009).
37. Pobric, G., Jefferies, E. & Lambon Ralph, M. A. Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci.* **104**, 20137–20141 (2007).
38. Pobric, G., Jefferies, E. & Lambon Ralph, M. A. Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia* **48**, 1336–1342 (2010).
39. Pobric, G., Jefferies, E. & Lambon Ralph, M. A. Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr. Biol.* **20**, 964–968 (2010).
40. Binder, J. R. *et al.* Toward a brain-based componential semantic representation. *Cogn. Neuropsychol.* **33**, 130–174 (2016).
41. Patterson, K., Nestor, P. J. & Rogers, T. T. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* **8**, 976–987 (2007).
42. Jackson, R. L., Rogers, T. T. & Lambon Ralph, M. A. Reverse-engineering the cortical architecture for controlled semantic cognition. *Nat. Hum. Behav.* **5**, 774–786 (2021).
43. Rogers, T. T. & McClelland, J. L. *Semantic Cognition: A Parallel Distributed Processing Approach*. (MIT Press, Cambridge, Mass., 2004).
44. Rogers, T. T. *et al.* Evidence for a deep, distributed and dynamic code for animacy in human ventral anterior temporal cortex. *eLife* **10**, e66276 (2021).
45. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, (2008).
46. Chen, Y. *et al.* The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex* **79**, 1–13 (2016).
47. Mahon, B. Z., Schwarzbach, J. & Caramazza, A. The representation of tools in left parietal cortex is independent of visual experience. *Psychol. Sci.* **21**, 764–771 (2010).
48. Tranel, D., Damasio, H. & Damasio, A. R. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* **35**, 1319–1327 (1997).
49. Chao, L. L., Haxby, J. V. & Martin, A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* **2**, 913–919 (1999).
50. Chen, L., Lambon Ralph, M. A. & Rogers, T. T. A unified model of human semantic knowledge and its disorders. *Nat. Hum. Behav.* **1**, 0039 (2017).
51. Price, A. R., Bonner, M. F., Peelle, J. E. & Grossman, M. Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *J. Neurosci.* **35**, 3276–3284 (2015).
52. Geschwind, N. Language and the brain. *Sci. Am.* **226**, 76–83 (1972).
53. Seghier, M. L., Fagan, E. & Price, C. J. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J. Neurosci.* **30**, 16809–16817 (2010).
54. Humphreys, G. F., Lambon Ralph, M. A. & Simons, J. S. A unifying account of angular gyrus contributions to episodic and semantic cognition. *Trends Neurosci.* **44**, 452–463 (2021).
55. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).
56. Dilkina, K. & Lambon Ralph, M. A. Conceptual structure within and between modalities. *Front. Hum. Neurosci.* **6**, (2012).

57. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* **59**, 2142–2154 (2012).
58. Afyouni, S. & Nichols, T. E. Insight and inference for DVARS. *NeuroImage* **172**, 291–312 (2018).
59. Frisby, S. L. *et al.* All spectral frequencies of neural activity reveal semantic representation in the human anterior ventral temporal cortex. 2025.04.17.649404 Preprint at https://doi.org/10.1101/2025.04.17.649404 (2025).
60. Clarke, A. Dynamic activity patterns in the anterior temporal lobe represents object semantics. *Cogn. Neurosci.* **11**, 111–121 (2020).
61. Rupp, K. *et al.* Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *NeuroImage* **148**, 318–329 (2017).
62. Wang, W., Degenhart, A. D., Sudre, G. P., Pomerleau, D. A. & Tyler-Kabara, E. C. Decoding semantic information from human electrocorticographic (ECoG) signals. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* **2011**, 6294–6298 (2011).
63. Sato, N. *et al.* Frequency-dependent cortical interactions during semantic processing: an electrocorticogram cross-spectrum analysis using a semantic space model. *Cereb. Cortex* **31**, 4329–4339 (2021).
64. Matoba, K. *et al.* Basal temporal language area revisited in Japanese language with a language function density map. *Cereb. Cortex* **34**, bhae218 (2024).
65. Cox, C. R. Testing neurocognitive predictions of the hub-and-spoke model of semantic memory with network representational similarity analysis. (University of Wisconsin-Madison, 2016).
66. Handjaras, G. *et al.* Modality-independent encoding of individual concepts in the left parietal cortex. *Neuropsychologia* **105**, 39–49 (2017).
67. Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J. & Binder, J. R. Decoding the information structure underlying the neural representation of concepts. *Proc. Natl. Acad. Sci.* **119**, e2108091119 (2022).
68. Humphreys, G. F. & Tibon, R. Dual-axes of functional organisation across lateral parietal cortex: the angular gyrus forms part of a multi-modal buffering system. *Brain Struct. Funct.* **228**, 341–352 (2023).
69. Giallanza, T., Campbell, D., Cohen, J. D. & Rogers, T. T. An integrated model of semantics and control. *Psychol. Rev.* (2024) doi:10.1037/rev0000485.
70. Morrison, C. M., Chappell, T. D. & Ellis, A. W. Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Q. J. Exp. Psychol. Sect. A* **50**, 528–559 (1997).
71. Barry, C., Morrison, C. M. & Ellis, A. W. Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Q. J. Exp. Psychol. Sect. A* **50**, 560–585 (1997).
72. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).
73. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).
74. Halchenko, Y. O. *et al.* HeuDiConv — flexible DICOM conversion into structured directory layouts. *J. Open Source Softw.* **9**, 5839 (2024).
75. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
76. Gorgolewski, K. J. *et al.* Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinformatics* **5**, (2011).
77. Markiewicz, C. J. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. Zenodo https://doi.org/10.5281/zenodo.12774221 (2024).
78. O'Brien, K. R. *et al.* Robust T1-weighted structural brain imaging and morphometry at 7T using MP2RAGE. *PLOS ONE* **9**, e99676 (2014).

79. Gaser, C. *et al.* CAT – A computational anatomy toolbox for the analysis of structural MRI data. 2022.06.11.495736 Preprint at https://doi.org/10.1101/2022.06.11.495736 (2023).
80. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
81. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
82. Avants, B. B., Tustison, N. & Johnson, H. Advanced Normalization Tools (ANTS). (2009).
83. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**, 2033–2044 (2011).
84. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
85. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed.* **10**, 171–178 (1997).
86. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* **20**, 870–888 (2003).
87. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
88. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
89. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219 (2004).
90. The tedana Community *et al.* ME-ICA/tedana: 23.0.1. Zenodo https://doi.org/10.5281/zenodo.7926293 (2023).
91. DuPre, E. *et al.* TE-dependent analysis of multi-echo fMRI with tedana. *J. Open Source Softw.* **6**, 3669 (2021).
92. Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M. & Bandettini, P. A. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage* **60**, 1759–1770 (2011).
93. Kundu, P. *et al.* Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc. Natl. Acad. Sci.* **110**, 16187–16192 (2013).
94. Posse, S. Multi-echo acquisition. *NeuroImage* **62**, 665–671 (2012).
95. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* **20**, 870–888 (2003).
96. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage* **9**, 179–194 (1999).
97. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage* **9**, 195–207 (1999).
98. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–289 (2002).
99. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, (2010).
100. Zhou, Y., Vales, M. I., Wang, A. & Zhang, Z. Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction. *Brief. Bioinform.* **18**, 744–753 (2017).
101. Stelzer, J., Chen, Y. & Turner, R. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage* **65**, 69–82 (2013).
102. Phipson, B. & Smyth, G. K. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, (2010).
103. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).