nature human behaviour

ARTICLES

PUBLISHED: 1 MARCH 2017 | VOLUME: 1 | ARTICLE NUMBER: 0039

A unified model of human semantic knowledge and its disorders

Lang Chen[1,2]*, Matthew A. Lambon Ralph[3]* and Timothy T. Rogers[1]*

How is knowledge about the meanings of words and objects represented in the human brain? Current theories embrace two radically different proposals: either distinct cortical systems have evolved to represent different kinds of things, or knowledge for all kinds is encoded within a single domain-general network. Neither view explains the full scope of relevant evidence from neuroimaging and neuropsychology. Here we propose that graded category-specificity emerges in some components of the semantic network through joint effects of learning and network connectivity. We test the proposal by measuring connectivity amongst cortical regions implicated in semantic representation, then simulating healthy and disordered semantic processing in a deep neural network whose architecture mirrors this structure. The resulting neuro-computational model explains the full complement of neuroimaging and patient evidence adduced in support of both domain-specific and domain-general approaches, reconciling long-standing disputes about the nature and origins of this uniquely human cognitive faculty.

Semantic memory supports the human ability to infer important but unobserved states of affairs in the world, such as object names ("That's a mushroom"), properties ("It is poisonous"), predictions ("It appears in autumn") and the meaning of statements ("It is edible after cooking"). Such inferences are generated within a cross-modal cortical network that encodes relationships amongst perceptual, motor and linguistic representations of objects, actions and statements (henceforth surface representations[1]). The large-scale architecture and organizational principles of the semantic network remain poorly understood, however. Theories about the nature and structure of this network have long been caught between two proposals: the system is modular and domain-specific, with components that have evolved to support different knowledge domains[2,3] (for example animals, tools, people); or it is interactive and domain-general, with all components contributing to all knowledge domains[4–6]. Despite profoundly different implications for the nature and roots of human cognition, these views have proven difficult to adjudicate[3,7].

We consider a third proposal, which arises from a general approach to functional specialization in the brain that we call connectivity-constrained cognition, or C^3 for convenience. This view proposes that functional specialization in the cortex is jointly caused by (1) learning/experience, (2) perceptual, linguistic and motor structures in the environment, and (3) anatomical connectivity in the brain. Connectivity is important because, within a given neurocognitive network, robustly connected components exert strong mutual influences and so, following learning, come to respond similarly to various inputs. In the case of semantic representation, these factors suggest a new approach that reconciles domain-specific and domain-general views. Specifically, learning, environmental structure and connectivity together produce graded domain-specificity in some network components because conceptual domains differ in the surface representations they engage[8–10]. For instance, tools engage praxis more than animals do[11], so regions that interact with action systems come to respond more to tool stimuli. Yet such effects emerge through domain-general learning of environmental structure, and centrally connected network components contribute critically to all semantic domains[12,13].

This C^3 proposal coheres with those of several other groups[8,14–17], but its potential to reconcile divergent views remains unclear because prior studies have focused on fairly specific questions about local network organization. Here we have tested the proposal by first measuring the anatomical connectivity of a broad cortical semantic network, and then assessing the consequences of that connectivity for healthy and disordered network behaviour using simulations with a deep neural network model. Specifically, from a new literature review and meta-analysis of functional brain imaging studies, we delineated cortical regions involved in semantic representation of words and visually presented objects, and identified those showing systematic semantic category effects. We then measured white-matter tracts connecting these regions using probabilistic diffusion-weighted tractography, resulting in a new characterization of cortical semantic network connectivity. From these results, we constructed a deep neural network model and trained it to associate surface representations of objects: their visual structure, associated functions and praxis, and words used to name or describe them. The resulting model is able to explain evidence adduced in support of both domain-specific and domain-general theories, including patterns of functional activation in brain imaging studies, impairments observed in the primary disorders of semantic representation, and the anatomical bases of these disorders.

Activation likelihood estimation (ALE) analysis

Prior empirical, modelling and neuroimaging work (see Supplementary Discussion 1) has identified several cortical regions that contribute to semantic processing and their respective functional roles, including: the posterior fusiform gyrus (pFG), which encodes visual representations of objects[18,19]; the superior temporal gyrus (STG), which encodes auditory representations of speech[20]; the lateral parietal cortex, which encodes representations of object

[1]Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, Wisconsin 53705, USA. [2]Stanford Cognitive and Systems Neuroscience Laboratory, 1070 Arastradero Road Suite 220, Palo Alto, California 94304, USA. [3]Neuroscience and Aphasia Research Unit (NARU), Division of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, M13 9PL, UK. *e-mail: lchen32@stanford.edu; matt.lambon-ralph@manchester.ac.uk; ttrogers@wisc.edu

function and praxis[19,21,22]; and the ventral anterior temporal lobe (ATL), thought to serve as a cross-modal hub that encodes semantic similarity structure[23,24]. To assess which of these regions show reliable semantic category sensitivity, and to identify additional category-sensitive regions not included among these, we conducted an ALE meta-analysis of functional imaging studies seeking semantic category effects. ALE provides a way of statistically assessing which category effects are reliably observed in the same location across studies. As in a prior meta-analysis[25], we included studies of activations generated by words or pictures denoting animals or artifacts (manmade objects). We identified 49 studies[9,19,21,26–71] with 73 independent experiments and 270 foci, making this the largest such analysis yet (for details, see Methods). Using recently updated ALE methods[72], we tested for cortical regions showing systematically different patterns for animals versus artifacts, or systematically elevated responses for both domains relative to baseline (see Supplementary Table 1). Results are shown in Fig. 1 and Supplementary Fig. 1.

Regions identified in prior work. Medial pFG is activated more for artifacts than animals bilaterally. Lateral pFG is activated above baseline for animals but not artifacts in both hemispheres, although the animal versus artifact contrast was only significant in the right, possibly because the homologous left-hemisphere region is 'sandwiched' between two areas showing the reverse pattern (posterior middle temporal gyrus (pMTG) and medial pFG; see Supplementary Fig. 2). The differential engagement of lateral/medial pFG by animal/artifact is well documented and typically thought to be bilateral[73].

STG did not show reliable category effects, consistent with the view from prior models[4,12,74] that it processes spoken word input and so should be equally engaged by animals and artifacts.

Ventral ATL did not exhibit activations above baseline for either domain, although this is not surprising for methodological reasons established in prior work[4,75]. Converging evidence from patient studies[76], brain imaging with appropriate methodology[4,75], transcranial magnetic stimulation[23], electro-corticography[77] and lesion-symptom mapping[24] have established the importance of ventral ATL for domain-general semantic processing. Prior models[5,12,78] included ventral ATL as a cross-modal semantic hub (see Supplementary Discussion 1.4 and 2.3).

Regions not specified or included in prior work. In the left parietal lobe, artifacts produced more activations than animals did, consistent with the proposal that this region encodes representations of object-directed action[19,79]. One prior model incorporated function representations in the lateral parietal cortex[12]. The cluster spanned the inferior and superior parietal lobes (IPL and SPL), which patient and imaging literatures suggest encode different aspects of action knowledge[79]. Thus we included both as separate regions of interest in the connectivity analysis and the model.

The pMTG exhibited more activations for artifacts than for animals, consistent with the literature implicating this region in the semantic representation of tools[25,73]. Accordingly, we included pMTG as a region of interest in the connectivity analysis and the model.

The lateral occipital complex (LOC) activated more often for animals than for artifacts, which probably reflects domain differences in visual structure, including greater complexity and more overlap among animals relative to manmade objects[80]. We thus identify LOC as a source of visual input to the inferotemporal cortex, and assume that animals generate more activation here because they have richer and more overlapping visual representations.

Semantic network connectivity

We next measured white-matter connectivity amongst all temporal regions of interest, and between temporal and parietal regions, using probabilistic diffusion-weighted tractography. We did not investigate intra-lobe connectivity within the parietal cortex[81,82], since these areas contribute to other non-semantic cognitive and perceptual abilities beyond the scope of this study. Diffusion-weighted images were collected from 24 participants using methods optimized to reduce the susceptibility artifact in ventral ATL[83]. Seeds were placed in the white matter underlying the regions of interest (ROI) from the ALE analysis or the literature (Figs 1 and 2; for ROI definition, see Methods), mapped back to native space for each subject (data and scripts are available on GitHub; please see Data Availability and Code Availability). The STG and LOC were excluded from the analysis since their connectivity is well-studied[84,85], and they are posited to provide spoken-word and visual input, respectively, to the semantic network. Results are shown in Fig. 2.

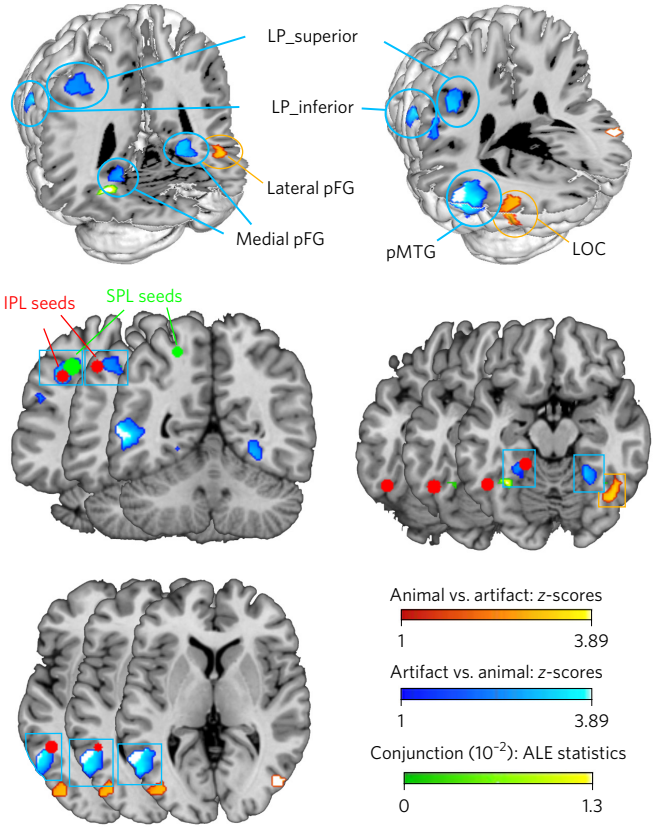

Figure 1 | ALE analysis showing regions that systematically respond more to animals than artifacts (orange), more to artifacts than animals (blue), or equally to both (green). Red dots indicate seed points from ALE analysis and literature review. LP, lateral parietal; LOC, lateral occipital complex.

Intratemporal connections. Both lateral and medial pFG projected into ATL (>5% in more than two-thirds of the participants) and to one another (Fig. 2a and Supplementary Table 2; for thresholding, see Methods). ATL also projected to both pFG regions and to the pMTG (>2.5% in more than half of the participants). Streamlines from pMTG terminated in the ATL neighbourhood (yellow stream in Fig. 2b) and projected to lateral pFG with high probability and to medial pFG with moderate probability (>1% in more than half of the participants).

Temporo-parietal connections. Streamlines from ATL did not extend into the parietal cortex, as also found previously[82,86]. Streamlines from pMTG, however, projected both to ATL and to IPL, providing an indirect route from IPL to ATL via pMTG (Fig. 2b). Likewise, the IPL streamlines projected to pMTG but not to ATL. Medial pFG did not stream to IPL, but did project more superiorly within the parietal lobe.

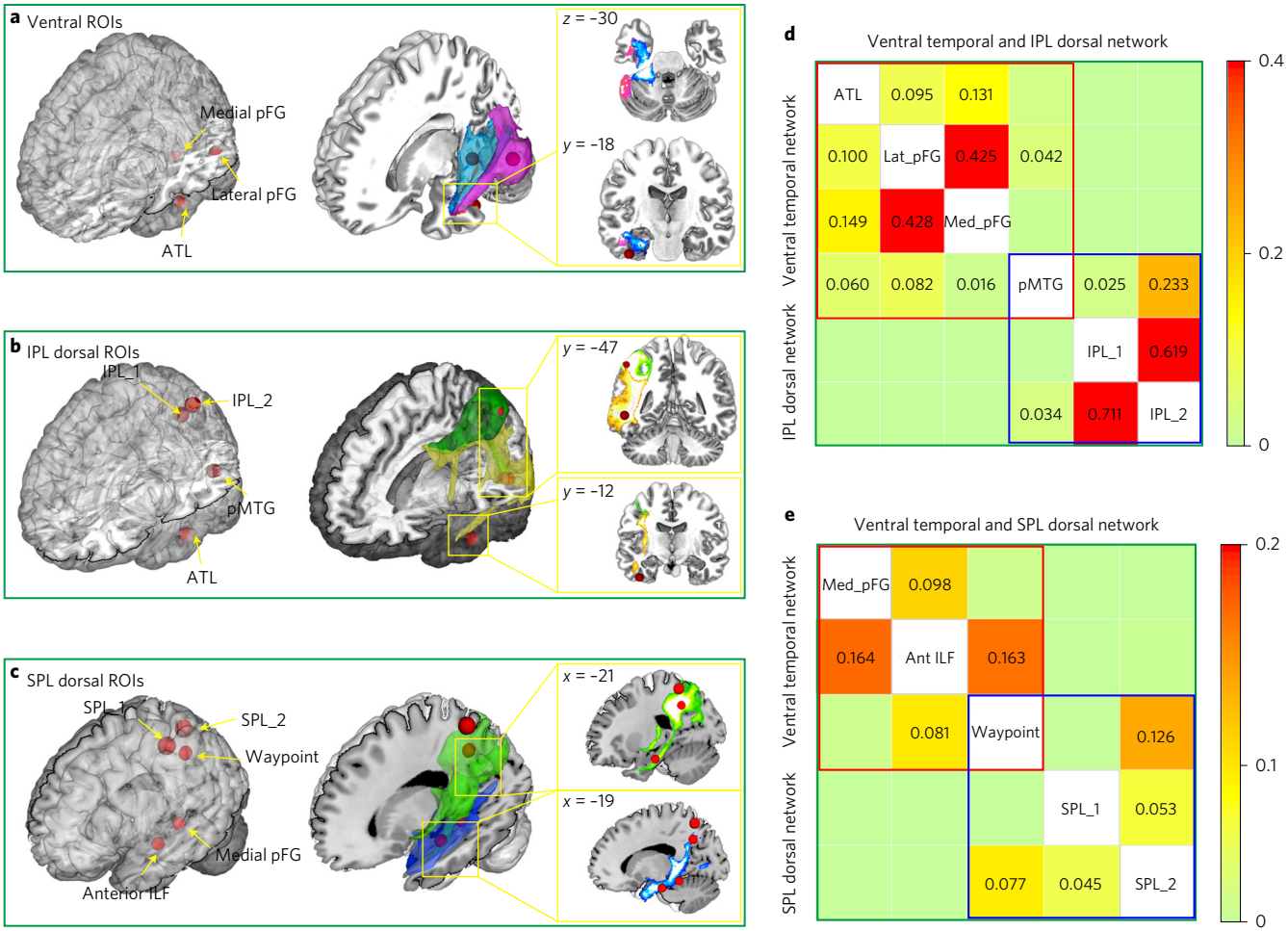

Figure 2 | Tractography results. Red spheres indicate seed points from ALE analysis and literature review. **a**–**c**, White-matter connectivity of cortical semantic network revealed by probablistic tractography. Streams from medial (blue) and lateral (pink) pFG project to ATL (**a**). Streams from pMTG (yellow) project to ATL and IPL, while IPL streams (green) project to pMTG but not ATL (**b**). Streams from inferior ATL white matter (blue) pass by medial pFG and branch superiorly, where they intersect SPL streamlines (green) (**c**). The waypoint seed was placed at this intersection. *x*, *y*, and *z* refer to coordinates in MNI (Montreal Neurological Institute) space. **d**,**e**, Matrices showing significant connectivity of temporal regions with IPL regions via the pMTG and with SPL regions via the tract identified by the waypoint seed. Numbers indicate group-averaged probability estimates (0–1) from seed (column) to target (row) regions.

Recent neuroanatomical studies from magnetic resonance tractography and tracing studies in non-human primates have suggested that the inferior longitudinal fasciculus (ILF), which connects ventral aspects of ATL, occipito-temporal and occipital cortex, also branches dorsally in its posterior extent to terminate in dorsoparietal regions[87,88] — potentially connecting ATL to SPL indirectly via the medial pFG. To test this possibility, we assessed the posterior trajectory of a seed more anteriorly along the ILF. The streamline passed through the medial pFG neighbourhood and branched superiorly into SPL (Fig. 2c). Likewise, SPL streamlines descended to intersect the ILF streamline. A waypoint seed placed at this junction streamed to SPL, the anterior ILF seed and medial pFG. Thus the tractography reveals two pathways from temporal to parietal regions of the network: one that connects ATL to IPL via the pMTG (Fig. 2d), and a second connecting ATL to the SPL via the medial pFG (Fig. 2e). This provides an *in vivo* demonstration of the dorsal-projecting ILF branch in humans.

An anatomically constrained computational model

Figure 3a shows a schematic of the ALE and connectivity results. We next constructed a neurocomputational model whose architecture mirrors these results, shown in Fig. 3b. The model is a deep recurrent neural network that computes mappings amongst visual representations of objects (coded in LOC), verbal descriptors (STG), and function (IPL) and praxis (SPL) action representations (see Methods and Supplementary Table 3). The model was trained with predictive error-driven learning to generate an item's full complement of visual, verbal, function and praxis properties, given a subset of these as input. Surface representations were generated to capture three well-documented aspects of environmental structure: (1) hierarchical similarity with few properties shared across domains, more shared within domains, and many shared within basic categories[11]; (2) many more praxis and function features for artifacts, and somewhat more visual features for animals[10,11]; and (3) more feature overlap amongst animals than artifacts[5] (see Supplementary Methods 5 and Supplementary Table 3).

We used the model to assess whether connectivity and learning jointly explain the category-specific patterns observed in the ALE meta-analysis. Fifteen models with different initial random weights were trained, providing analogues of 15 subjects in a brain imaging study. Models were tested with simulations of both word and picture comprehension. The activation patterns generated by these inputs were treated as analogues of the BOLD response and analysed to identify model regions showing systematic category effects[12] (see Methods).

Results are shown in Fig. 3c. All category effects observed in the ALE analysis emerged in the corresponding model layers for both words and pictures. Medial pFG, pMTG, IPL and SPL responded

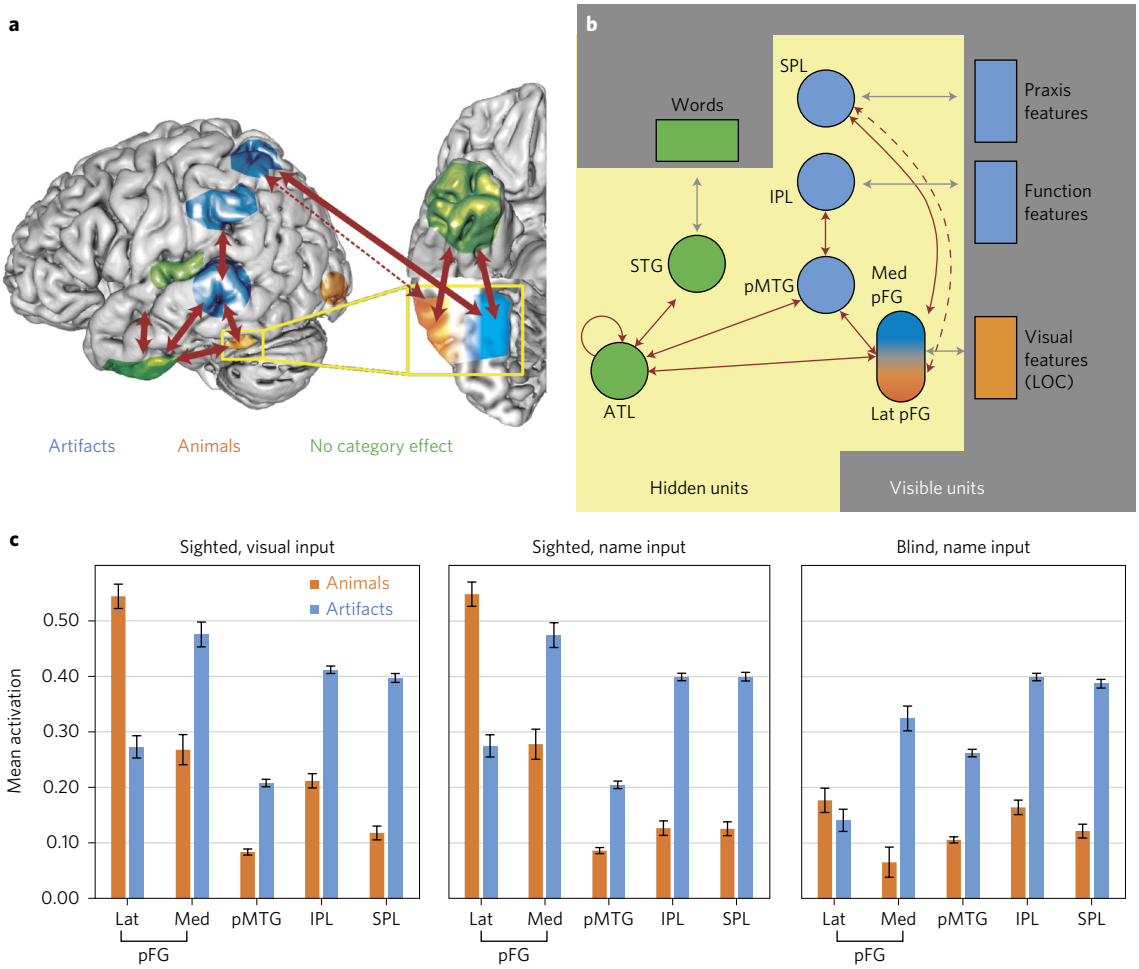

Figure 3 | Model architecture and functional magnetic resonance imaging (fMRI) data simulations. a, Schematic showing ALE and connectivity results. Red arrows indicate significant connectivity in tractography while coloured regions indicate semantic category effects in the ALE analysis. The dotted arrow indicates that connectivity diminishes from medial to lateral pFG. **b**, Architecture of the corresponding neural network model. Boxes indicate layers that directly encode features of objects (visible units) and circles indicate model analogues of cortical ROIs where representations are learned (hidden units). For visible units, blue indicates more active features for animals than artifacts while orange indicates the reverse. For hidden units, circle colour indicates expected category effects using the same scheme as panel **a**. Red arrows indicate model connections that correspond to tractography results; grey arrows indicate connections that mediate activation between visible and hidden units. **c**, Mean unit activation for animals or artifacts in each model ROI, for visual and word inputs of the 'sighted' model (left and middle) and for word inputs in the 'blind' model (right). Error bars represent standard errors of means across individual simulation runs.

more to artifacts because they strongly interact with function or praxis representations. Lateral pFG responded more to animals because the medial units had partially 'specialized' to represent artifacts. Thus model connectivity, learning and environmental structure together produced the category-sensitive activations observed in the ALE analysis.

Category-specific activations have also been observed during word comprehension in congenitally blind participants, providing important support for domain-specific views since such results cannot arise from domain differences in visual structure[9,63,89]. To assess whether learning and connectivity also explain such patterns, we replicated the simulations in models trained without visual inputs or targets. The animal advantage in lateral pFG disappeared, presumably because these units no longer communicate activation from early vision[12]. Artifacts, however, continued to elicit greater activation in medial pFG, posterior pMTG, IPL and SPL, because these units continue to participate in generating function and praxis representations for object-directed action. The absence of a category effect in lateral pFG and tool/praxis-specific activation patterns in pMTG, IPL and SPL have all been reported in this population[9,63,89].

Disorders of semantic representation

We next considered whether learning and connectivity explain the primary disorders of semantic representation and their anatomical basis. By 'primary', we mean acquired disorders that reflect degraded semantic representation rather than access/retrieval deficits[90,91] and also have been shown in case-series studies to manifest predictable patterns of impairment. These include: semantic dementia (SD), where progressive bilateral ATL atrophy produces a category-general semantic impairment[78]; herpes simplex viral encephalitis (HSVE), in which acute bilateral ATL pathology produces chronic impairments disproportionately affecting animals[92]; temporo-parietal tumour resection (TPT), which produces greater impairment for artifacts[93]; and forms of visual agnosia (VA) producing slower and less accurate recognition for animals[80,94].

Both SD and HSVE were simulated by removing increasing proportions of ATL connections. To capture the progressive nature of SD, performance was assessed without relearning after connections were removed. For HSVE, we considered two damage models. In the homogeneous variant (HSVE), damage was identical to SD, but the network was then retrained to simulate acute injury with

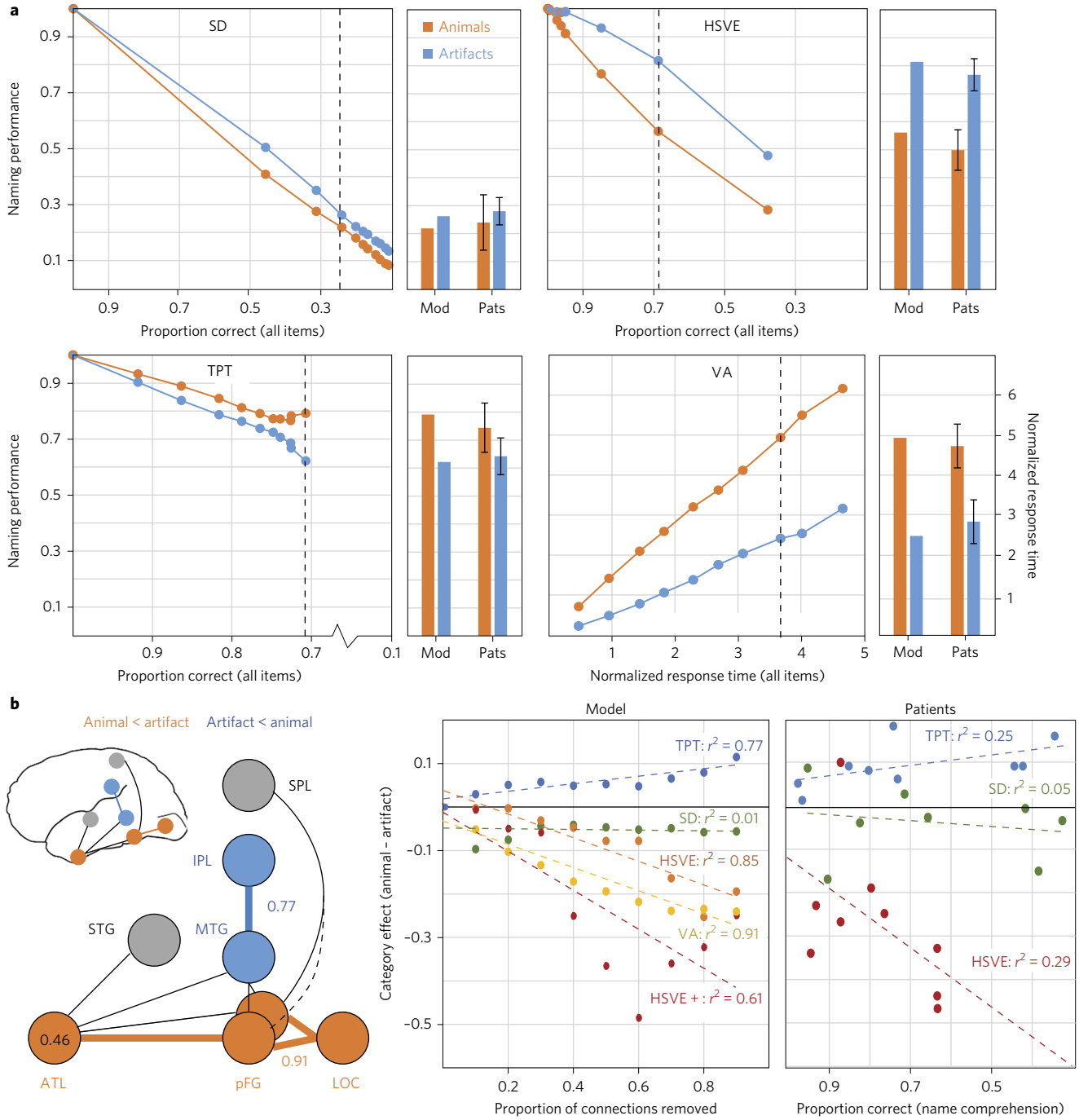

Figure 4 | Results of patient simulations. a, Line plots show model naming accuracy for animals and artifacts at 10 increasing levels of damage for each disorder plotted against overall accuracy (all items). Dashed vertical lines indicate the damage level that most closely matches mean overall accuracy in the corresponding patient group. HSVE data are for the homogeneous damage model (HSVE); data for the asymmetric damage model (HSVE+) appear in Supplementary Fig. 5. Bar plots show accuracy by category for the model (Mod) at this level compared with patient (Pats) means/standard errors reported elsewhere[78,93,94]. **b**, Lesion–symptom mapping results. Left, layers/connections in which lesion size predicts increasing artifact (blue) or animal (orange) disadvantage. The dashed line represents weaker connectivity from lateral pFG to SPL. Middle, correlation between lesion size and category effect in each simulated patient group. Right, category effect size in naming plotted against overall impairment as measured by word comprehension (SD and HSVE) or overall naming (TPT) in case-series studies of real patients.

recovery. In the asymmetric variant (HSVE+), lateral connections between ATL and pFG were more likely to be removed than medial connections, consistent with a possible difference noted in a direct comparison of white-matter pathology in SD versus HSVE[92] (see Supplementary Discussion 3.2). The damaged model was again retrained. TPT was simulated by removing a proportion of connections within/between pMTG and IPL layers, while VA was simulated by removing weights between LOC and pFG layers. At each level of damage for each disorder, we simulated picture naming for all animal and artifact items (see Data Availability and Code Availability).

Results are shown in Fig. 4. The model captures the direction and magnitude of several key phenomena including: no category effect in SD; a substantial animal disadvantage in both HSVE variants

(results of HSVE+ in Supplementary Fig. 5); a modest artifact disadvantage in TPT; an animal disadvantage in response time in VA; worse anomia in SD than HSVE; and a smaller and opposite category effect in TPT compared with HSVE.

The pattern of network connectivity transparently explains the key results for two patient groups: TPT, where disrupted interactions with function representations in IPL disproportionately affect artifacts, and SD, where ATL damage produces a domain-general impairment. In VA, the category effect arises from 'visual crowding'[80]: because animals overlap more in their visual properties[5,11], they are more difficult to discriminate (and hence to name) when inputs from vision are impoverished[95]. In HSVE, the model pathology is identical to SD; the category effect thus arises through re-learning, by means of two mechanisms. First, intact function and praxis layers can support new learning for items with these properties, that is, for artifacts. Second, because animals share more properties than artifacts do, their ATL representations are also 'conceptually crowded', compromising relearning of inter-item differences when ATL representations are damaged[78]. In HSVE+, the effect is magnified when lateral ATL–pFG connections are disproportionately removed, since these connections provide more support for animal knowledge, as shown in the simulation of imaging results (see Supplementary Figs 4 and 5).

The simulation suggests a new resolution to the long-standing puzzle of why patients with HSVE and SD show qualitatively distinct impairments despite largely overlapping pathology[78]. Category-specific deficits may arise when white-matter pathology is distributed asymmetrically in the ATL, but even when pathology is identical they may emerge through relearning following the acute injury (see Supplementary Discussion 3.2 and 4). To assess this hypothesis, we first evaluated model predictions by regressing the magnitude of the category effect (artifact accuracy minus animal accuracy) on the total amount of damage, the amount of relearning and their interaction, in the simulation of both HSVE and HSVE+ (see details in Supplementary Methods 6). In both cases the two factors interacted reliably: when damage was severe, relearning produced a larger category effect, but when damage was mild, relearning shrank the category effect (Supplementary Fig. 7A and B; interaction for HSVE $t = 2.501$, $P = 0.014$; interaction term for HSVE+ $t = 2.137$, $P = 0.035$). We then assessed whether the same pattern is observed in the literature. Across 19 previously reported HSVE cases of category-specific impairment (Supplementary Table 6), we regressed the reported category effect on the overall severity of the impairment, the amount of relearning (assessed as the time elapsed between injury and test), and their interaction term. Consistent with model predictions, these factors interacted reliably ($t = 3.298$, $P < 0.01$): relearning produced larger effects when deficits were severe but smaller effects when deficits were mild (Supplementary Fig. 7C). The same pattern was also observed longitudinally in four patients with HSVE[96,97] (Supplementary Fig. 7D). By contrast, this pattern was not found in the non-HSVE cases (for full results, see Supplementary Table 7). Thus the model's account of category-specific impairment is consistent with the existing literature.

Finally, we considered classic lesion-symptom mapping results suggesting that animal-selective deficits occur with ventro-temporal damage whereas artifact-selective deficits occur with temporo-parietal pathology[98]. We conducted a model lesion-symptom analysis by grouping simulated patients across all four disorders into a single dataset. We quantified regional pathology in every model patient as the proportion of connections removed from each layer and measured category selectivity as the difference in accuracy naming artifacts versus animals. We then computed, across all patients at each layer, the correlation between pathology and category selectivity.

Figure 4b shows the results. Damage in ventral temporal model regions (ATL, pFG and LOC) significantly predicted greater impairment for animals than artifacts, whereas damage in pMTG and IPL regions predicted the reverse pattern (Fig. 4b, left). Importantly, the ATL effect was only carried by the HSVE simulations: SD simulations alone showed no relationship between lesion severity and category effect (Fig. 4b, middle). The same pattern is observed in case-series studies of the corresponding syndromes for which data is available (Fig. 4b, right). Thus the model explains both the canonical lesion-symptom results and their puzzling discrepancy with SD.

Discussion

We have proposed a new neurocomputational model for the neural bases of semantic representation which, in building on contributions from several groups[19,93,99], unifies domain-specific and domain-general approaches. The core and critical theoretical contribution is that initial connectivity, domain-general learning and environmental structure all jointly shape functional activation within the cortical semantic network, leading to graded category-specificity in some network components but domain-general processing in the ATL hub, within a network whose principal function is to support cross-modal inference. The model explains the neuroimaging and patient phenomena central to both domain-specific and domain-general theories, including (1) category-specific patterns of functional activation in sighted and congenitally blind individuals, (2) patterns of impairment observed across four different neuropsychological syndromes, and (3) the anatomical bases of these patterns. It also exemplifies a general approach to functional specialization in cortex that we have termed connectivity-constrained cognition or C^3.

Our model reconciles and extends several competing perspectives in the literature. Like the sensory-functional hypothesis, category sensitivity arises from domain differences in the recruitment of action versus visual representations[10,100]; but we show that learning and connectivity can produce domain differences even in the absence of visual experience, and outside canonical action areas, addressing key criticisms of the sensory-functional view[3]. Like the distributed domain-specific hypothesis, category-sensitivity reflects network connectivity, with temporo-parietal pathways initially configured to assist vision–action relationships important for tool knowledge[9]. The model reconciles this perspective with the extensive evidence for domain-general representation in the ATL. An important account of optic aphasia relied on graded functional specialization arising from constraints on local connectivity[8]; our model extends this idea to incorporate long-range connectivity constraints. As in the correlated-structure view, category-selectivity arises partly from different patterns of overlap among animal versus artifact properties[6], but in our model, connectivity also plays a critical role. Finally, this work extends the hub-and-spoke model under which the ATL constitutes a domain-general semantic hub for computing mappings amongst all surface modalities[4]. The model illustrates how domain-specific patterns can arise within the 'spokes' of such a network, even while the ATL plays a critical domain-general role in semantic representation[13] (see Supplementary Discussion 2 for relationship to other models).

In emphasizing semantic representation, we have not considered the fronto-parietal systems involved in semantic control[101], nor does the model address open questions about lateralization, abstract and social concepts, or other conceptual distinctions amongst concrete objects. We therefore view the proposed model as establishing a foundation rather than an end point. Nevertheless, the current work is unique in developing a neurocognitive model whose architecture is fully constrained, a priori, by systems-level neural data. The project illustrates how simulation models at this level of abstraction can provide an important conceptual bridge for relating structural and functional brain imaging to healthy and disordered cognitive functioning[74]. Although interest in neural networks has recently rekindled in machine learning[102], their original promise as tools for bridging minds and brains[103] has remained largely untested. We have

shown that the convergent use of network simulation models with the other tools of cognitive neuroscience can produce new insights with the potential to resolve otherwise pernicious theoretical disputes. We further believe that the C^3 approach we have sketched, in which network models are used to illuminate how connectivity, learning and environmental structure give joint rise to cognitive function, can be similarly useful in other cognitive domains.

Methods

ALE analysis. We followed the standard literature search procedure from previous ALE studies[25,104] and found 49 papers describing 73 independent studies (31 for animal and 42 for artifact; for study selection, see Supplementary Methods 1) up to July 2013 and reporting a total of 270 foci (103 for animal and 167 for artifact). The ALE meta-analysis was carried out with the software package gingerALE v2.3[105,106]. The ALE analysis strictly followed the steps proposed by Price *et al.*[104] and Eickhoff *et al.*[72,105,107], and coordinates in MNI (Montreal Neurological Institute) space were used for ALE analysis and reports. Main effects of animal and artefact concepts (concordance of foci showing greater activations for animal versus baseline and artefact versus baseline) are reported in Supplementary Table 1. Next, we combined the resulting ALE animal and artifact maps and tested for brain regions commonly activated by both categories (conjunction analysis) and showing reliably different activations for the two categories of interest (contrast analysis) as reported in the main text.

Connectivity analysis. Diffusion-weighted images were collected from 24 right-handed healthy subjects (11 female; mean age = 25.9) at the University of Manchester, UK[86]. All participants are right-handed as determined by the Edinburgh Handedness Inventory[106]. Inclusion and exclusion criteria were stated in previous studies[86,108], and no randomization or blinding was needed. Informed consents were obtained for all subjects.

Image acquisition. Imaging data were acquired on a 3T Philips Achieva scanner (Philips Medical Systems, Best, Netherlands), using an eight-element SENSE head coil. Diffusion-weighted imaging was performed using a pulsed gradient spin echo echo-planar sequence with TE = 59 ms, TR ≈ 1,500 ms (cardiac gated), $G = 62\,\mathrm{mT\,m^{-1}}$, half-scan factor = 0.679, 112 × 112 image matrix reconstructed to 128 × 128 using zero padding, reconstructed resolution 1.875 × 1.875 mm, slice thickness 2.1 mm, 60 contiguous slices, 61 non-collinear diffusion sensitization directions at b-value = 1,200 s mm^{-2} ($\Delta = 29.8$ ms, $\delta = 13.1$ ms), 1 at $b = 0$, and SENSE acceleration factor 2.5. A high-resolution T1-weighted 3D turbo field echo inversion recovery scan (TR ≈ 2,000 ms, TE = 3.9 ms, TI = 1,150 ms, flip angle 8°, 256 × 205 matrix reconstructed to 256 × 256, reconstructed resolution 0.938 × 0.938 mm, slice thickness 0.9 mm, 160 slices, SENSE factor 2.5) was also acquired for the purpose of high-precision anatomical localization of seed regions for tracking. Distortion correction to remediate signal loss in ventral ATL was applied, using the same method reported in other studies[86,108].

ROI definition. The ROIs were chosen to reside in the white-matter underlying the peaks identified in the ALE-meta analysis, or from regions reported in the relevant literature. Specifically, ROIs in lateral pFG, medial pFG, MOG, pMTG, IPL (IPL_1) and SPL (SPL_1) in the left hemisphere were chosen from the ALE meta-analysis as regions showing reliable category-specific activation patterns. The ATL ROI was chosen from a functional MRI study[109] that reported cross-modal activation for conceptual processing in the ATL. Owing to the uncertainty of tempo-parietal connectivity, we also included a second IPL seed (IPL_2) whose coordinates were chosen from a study in which TMS (transcranial magnetic stimulation) to this region slowed naming of tools but not animals[110]. Likewise we included a second SPL ROI (SPL_2) reported by Mahon *et al.*[63] as a peak showing preferential activation for artifact stimuli in both sighted and congenitally blind participants. To assess the caudal-going trajectory of the ILF, we placed an additional seed in the inferior temporal white matter at the anteriormost extent of the artifact peak revealed by the ALE meta-analysis. As reported in the main text, this streamline branched superiorly up into parietal cortex, intersecting the streamline from the SPL seeds. To determine whether a single tract might connect SPL, medial pFG and ATL, we placed a final waypoint seed at this intersection. For more details about ROI definitions, see Supplementary Methods 2.

Probabilistic tracking procedure. We restricted our analysis to the left hemisphere, and following similar procedure to a previous study[108], a sphere with a diameter of 6 mm centred on the seed coordinate for each ROI was then drawn in the MNI template (see Supplementary Table 2 for the exact coordinates; details in Supplementary Information). Finally, the ROIs defined in a common space were converted into the native brain space of each individual.

For each voxel within a seed ROI sphere, 15,000 streamlines were initiated for unconstrained probabilistic tractography using the probabilistic index of connectivity (PICo) method[108,111]. The step size was set to 0.50 mm. Stopping criteria for the streamlines were set so that tracking terminated if the pathway curvature over a voxel was greater than 180, or the streamline reached a physical path limit of 500 mm. In the native-space tracking data from each seed region for each individual, ROI masks were overlaid, and a maximum connectivity value (ranging from 0 to 15,000) was obtained for the seed region and each of the other ROIs, resulting in a matrix of streamline-based connectivity. A standard two-level threshold approach was applied to determine high likelihood of connection in this matrix[108]. At each individual level, three thresholds, 1% (lenient), 2.5% (standard) and 5% (stringent), were used to investigate the probable tracts in a wider range. At the group level, only connections present in at least half (≥12) subjects were considered highly probable across subjects (for more details of thresholding, see Supplementary Methods 3). A group-averaged tractography image was then obtained by averaging the normalized individual data[108].

Computer simulations of functional MRI data. The model architecture shown in Fig. 3b (main text) was implemented using the Light Efficient Network Simulator (LENS) software[112]. The model included four visible layers directly encoding model analogues of visual, verbal (names and descriptions), praxis and function properties of objects. Each visible layer was reciprocally connected with its own modality-specific hidden layer, providing model analogues to the pFG (visual hidden units), STG (verbal hidden units), IPL (function hidden units) and SPL (praxis hidden units). The model also included two further hidden layers corresponding to the ventral ATL and the posterior MTG. Hidden layers were connected with bidirectional connections matching the results of the tractography analysis, as shown in Fig. 3b. A spatial gradient of learning rate on visuo-praxic connections of units in the pFG layer along an anatomical lateral-to-medial axis was implemented[12] (for details, see Supplementary Methods 4), to capture the observation that medial pFG is more strongly connected to parietal regions than is lateral pFG[19]. All units used a sigmoidal activation function and were given a fixed bias of -2 so that, in the absence of input from other units, they tended to adopt a low activation state. Units updated their activation states continuously using a time integration constant of 30. Model implementation and training environment files can be downloaded online (see Data Availability).

Training environment. A model environment was constructed to contain visual, verbal, function/action and praxis representations for 24 different exemplars of animals and 24 different exemplars of tools, with each domain organized into four basic categories, each containing six exemplars (for representational schemes of training exemplars, see Supplementary Table 3 and Supplementary Methods 5). In total, there were 48 training exemplars. Visual and verbal representations for each item in this set were generated stochastically in accordance with the constraints identified by Rogers *et al.*[5] in their analysis of verbal attribute-listing norms and line drawings of objects. Thus, items in different domains shared few properties; items within the same category shared many properties; animals from different categories shared more properties than did artifacts from different categories; and animals had more properties overall than did artifacts. Each item was also given a unique name as well as a label common to all items in the same category.

Praxis representations were also constructed for each item, taking the form of distributed patterns over the 10 units in the visible praxis layer[12]. For all animal items, these units were turned off. For artifacts, distributed patterns were created that covaried with, but were not identical to, the item's corresponding visual pattern, as a model analogue of vision-to-action affordances. Function representations simply duplicated the praxis patterns across the 10 visible units for function features.

Model training procedures. The model was trained to generate, given partial information about an item as input, all of the item's associated properties, including its name, verbal description, visual, function and praxis features, similar to our previous work[12] (for details, see Supplementary Methods 5). Weights were updated using a variant of the backpropagation learning algorithm suited to recurrent neural networks, using a base learning rate of 0.01 and a weight decay of 0.0005 without momentum[113]. 'Congenitally blind' model variants were trained with the same parameters on the same patterns, but without visual experience: visual inputs were never applied to the model, and visual units were never given targets. All models were trained exhaustively for 100k epochs at which point they generated correct output (for details, see Supplementary Methods 5) across all visible units for the great majority (>94%) of inputs. For each model population (sighted/blind), 15 different subjects were simulated with different model training runs, each initialized with a different set of weights sampled from a uniform random distribution with mean 0 and range ±0.1 (for model performance after training, see Supplementary Table 4).

Simulating functional brain imaging studies. The brain imaging studies simulated involved two tasks: picture viewing, in which participants made a semantic judgement from a picture of a familiar item, and name comprehension, in which they made a semantic judgement from the spoken name of a familiar item. To simulate the picture viewing task in sighted model variants, the visual feature pattern corresponding to a familiar item was applied to visual input units and the trained model cycled until it reached a steady state. To simulate name comprehension in both sighted and congenitally blind variants, a single unit corresponding to the item's name was given excitatory external input, and the

model again cycled until it reached a steady state. In both tasks, after settling, the activation of each model unit was recorded and taken as an analogue of the mean activity difference from baseline for a population of neurons at a single voxel. This value was then distorted with Gaussian noise ($\mu = 0$, $\sigma^2 = 0.1$) to reflect the error in signal estimation intrinsic to brain imaging methods. The response of each unit was then averaged across items in each condition (animal versus artifact) and then spatially smoothed with a Gaussian kernel ($\mu = 0$, $\sigma^2 = 1$) encompassing two adjacent units. A group-level contrast was performed to find the peak activation for both animal and artifact concepts using the averaged data across the 15 model subjects. ROI analysis was then performed on activation value of the peak unit averaged together with two neighbouring units on either side (four in total).

Computer simulations of patient data. Following simulation of functional imaging data, we assessed whether the model could explain patterns of impaired semantic cognition and their neuroanatomical basis in four disorders of semantic representation. Here, we provide basic information of the phenotype of each disorder and model simulation procedure (details of pathology and motivation are given in Supplementary Discussion 3). The model architecture and training environment were the same as in the simulations of brain imaging data, except that pattern frequencies were adapted to ensure that the names of animal and artifact items appeared as inputs and targets with equal frequency (see Supplementary Fig. 3).

(1) SD is a neurodegenerative disorder associated with gradual thinning of cortical grey matter and associated white-matter fibres, centred in the ATL[24], and produces a robust, progressive and yet selective deterioration of semantic knowledge for all kinds of concepts, across all modalities of reception and expression[5,76,114]. We simulated SD by removing an increasing proportion of all weights entering, leaving or internal to the ATL hidden layer uniformly from 0.1 to 1.0 with an increment of 0.1. At each level of damage, the model was tested without allowing it to relearn or reorganize.

(2) HSVE is a disease that produces rapid bilateral necrosis of grey and white matter, generally encompassing the same regions affected in SD, but patients with semantic impairments from HSVE have been found with greater damage in temporal white matter especially in the lateral axis compared with SD[92]. HSVE patients often show less semantic impairment overall, with greater deficits of knowledge for animals than for manmade objects[78,92].

We consider two potential explanations, each associated with a different model of HSVE pathology. The model first captures differences in the time-course of SD versus HSVE: whereas the former progresses slowly over a course of years, the latter develops rapidly and is then halted by antiviral medication, after which patients often show at least some recovery of function. Weights were removed from the ATL layer as in the SD simulation, but the damaged model was then retrained before assessment on the naming task (for motivation, see Supplementary Discussion 3.2). Retraining used the same parameters used in the last cycle of the initial training, namely, learning rate 10^{-3} and weight decay 10^{-6}. The main text reports data following 3k epochs of retraining when the model performance had largely stabilized (see Supplementary Fig. 4 for recovery trajectory). The second further assessed the potential contribution of differential white-matter damage across the lateral/medial axis of the ATL in HSVE[92]. To simulate this, a proportion of ATL connections selecting uniformly with probability p were removed in a first pass (as in the SD simulation), then a second removal of connections was applied to weights between ATL and lateral pFG units (units 0–9). In a 30% lesion, for instance, 30% of all ATL connections entering or leaving each ATL unit were removed, and then 30 % of the original connections between ATL and lateral pFG units were additionally removed. Thus, when the global lesion severity equaled or exceeded 50%, all connections between ATL and lateral pFG were removed. Finally, the model was retrained as in the homogeneous variant of HSVE, and performance on the retrained model was assessed (see Supplementary Fig. 5). We also demonstrated that without relearning, the HSVE variant showed little evidence for category-specific impairment (just as with SD simulations), but the HSVE+ variant showed more severe impairment in the animal category (see Supplementary Fig. 6).

(3) TPT. Campanella and colleagues[93] presented the first relatively large-scale case-series study of artifact-category impairment in a group of 30 patients who had undergone surgical removal of temporal-lobe tumours. The group exhibited significantly worse knowledge of nonliving things than of animals, with difference scores in naming accuracy ranging from 2% to 21%. Voxel-based lesion-symptom mapping revealed that the magnitude of the category effect was predicted by pathology in posterior MTG, inferior parietal cortex, and the underlying white matter. To simulate this pathology in the model, we removed connections between and within IPL and MTG model regions uniformly from 0.1 to 1.0 with an increment of 0.1.

(4) Category-specific VA. Finally, a long tradition of research suggests that forms of associative VA arising from damage to occipitotemporal regions can have a greater impact on recognition of living than non-living things[115,116]. The deficit is specific to vision, and more evident in naming latency rather than accuracy at milder impairment[94]. To capture disordered visual perception, we removed a proportion of the weights projecting from the visual input layer (LOC) to the visual hidden layer (pFG). We used a smaller range from 0.025 to 0.25 in increments of 0.025 to preserve sufficient visual inputs to the system.

Assessment of model performance. For each disorder, model performance was assessed on simulated picture naming. For each item, the corresponding visual features were given positive input, and the activations subsequently generated over units encoding basic-level names were inspected to assess performance. Naming performance was scored as correct if the target name unit was both the most active of all basic name units and activated above 0.5; otherwise it was scored as incorrect. In VA at mild impairment, the category-specific impairment can be observed in response time so that we computed naming latency as the number of update cycles (ticks) required for the target unit to reach an activation of 0.5 (for correct naming trials only). For comparison to standardized human latency data in VA, the model latency was standardized by computing $(N_j - N_0)/N_0$, where N_j is the number of ticks used for the model to produce a response at the jth level of lesion severity and N_0 the number of ticks for naming without any lesion in the model[117]. Therefore, the raw latency measure is adjusted by baseline response latency differences between categories that exist in the performance of the intact models. Note that in figures of the main text, the severity was recomputed as the overall naming accuracy, collapsing the animal and artifact categories. See Supplementary Table 5 for naming accuracy and latency at different levels of lesion severity measured as percentage of affected connections.

Code availability. Program scripts to run the tractography analysis and model simulations, including model structure specification, training and testing procedure, and lesion procedure are available on GitHub at https://github.com/halleycl/ChenETAL_NatHumanBehav_SI-Online-materials.

Data availability. Source data include coordinates used in the meta-analysis, computed tractography connectivity index, model initial weights and train weights, and training environment. Output data include the output files from the meta-analysis, individual connectivity matrix, model activation and output files, and summary data files of all simulations. These data are available on GitHub at https://github.com/halleycl/ChenETAL_NatHumanBehav_SI-Online-materials.

Received 27 June 2016; accepted 15 December 2016; published 1 March 2017

References

1. Fernandino, L. *et al.* Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia* **76,** 17–26 (2015).
2. Caramazza, A. & Shelton, J. R. Domain-specific knowledge systems in the brain: the animate–inanimate distinction. *J. Cogn. Neurosci.* **10,** 1–34 (1998).
3. Caramazza, A. & Mahon, B. Z. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn. Sci.* **7,** 354–361 (2003).
4. 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).
5. Rogers, T. T. *et al.* The structure and deterioration of semantic memory: a computational and neuropsychological investigation. *Psychol. Rev.* **111,** 205–235 (2004).
6. Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R. & Levy, J. P. Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain Lang.* **75,** 195–231 (2000).
7. Chen, L. & Rogers, T. T. Revisiting domain-general accounts of category specificity in mind and brain. *Wiley Interdiscip. Rev. Cogn. Sci.* **5,** 327–44 (2014).
8. Plaut, D. C. Graded modality-specific specialisation in semantics: a computational account of optic aphasia. *Cogn. Neuropsychol.* **19,** 603–639 (2002).
9. Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M. & Caramazza, A. Category-specific organization in the human brain does not require visual experience. *Neuron* **63,** 397–405 (2009).
10. Warrington, E. K. & Shallice, T. Category specific semantic impairments. *Brain* **107,** 829–854 (1984).
11. Cree, G. S. & McRae, K. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J. Exp. Psychol. Gen.* **132,** 163–201 (2003).
12. Chen, L. & Rogers, T. T. A model of emergent category-specific activation in the posterior fusiform gyrus of sighted and congenitally blind populations. *J. Cogn. Neurosci.* **27,** 1981–1999 (2015).
13. 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).
14. Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512,** 423–426 (2014).
15. Gomez, J. *et al.* Functionally defined white matter reveals segregated pathways in human ventral temporal cortex associated with category-specific processing. *Neuron* **85,** 216–227 (2015).

16. Mahon, B. Z. & Caramazza, A. What drives the organization of object knowledge in the brain? *Trends Cogn. Sci.* **15,** 97–103 (2011).
17. Plaut, D. C. & Behrmann, M. Complementary neural representations for faces and words: a computational exploration. *Cogn. Neuropsychol.* **28,** 251–275 (2011).
18. Martin, A. & Chao, L. L. Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* **11,** 194–201 (2001).
19. Mahon, B. Z. *et al.* Action-related properties shape object representations in the ventral stream. *Neuron* **55,** 507–520 (2007).
20. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8,** 393–402 (2007).
21. Kellenbach, M. L., Brett, M. & Patterson, K. Actions speak louder than functions: the importance of manipulability and action in tool representation. *J. Cogn. Neurosci.* **15,** 30–46 (2003).
22. Chouinard, P. A & Goodale, M. A. Category-specific neural processing for naming pictures of animals and naming pictures of tools: an ALE meta-analysis. *Neuropsychologia* **48,** 409–418 (2010).
23. 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. USA* **104,** 20137–20141 (2007).
24. Acosta-Cabronero, J. *et al.* Atrophy, hypometabolism and white matter abnormalities in semantic dementia tell a coherent story. *Brain* **134,** 2025–2035 (2011).
25. Chouinard, P. A. & Goodale, M. A. Category-specific neural processing for naming pictures of animals and naming pictures of tools: an ALE meta-analysis. *Neuropsychologia* **48,** 409 (2010).
26. Hwang, K. *et al.* Category-specific activations during word generation reflect experiential sensorimotor modalities. *Neuroimage* **48,** 717–725 (2009).
27. Smith, C. D. *et al.* Differences in functional magnetic resonance imaging activation by category in a visual confrontation naming task. *J. Neuroimaging* **11,** 165–170 (2001).
28. Grossman, M. *et al.* The neural basis for category-specific knowledge: an fMRI study. *Neuroimage* **15,** 936–948 (2002).
29. Martin, A., Haxby, J. V, Lalonde, F. M., Wiggs, C. L. & Ungerleider, L. G. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* **270,** 102–105 (1995).
30. Tyler, L. K. *et al.* Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system. *Cogn. Neuropsychol.* **20,** 541–559 (2003).
31. Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D. & Damasio, A. R. A neural basis for lexical retrieval. *Nature* **380,** 499–505 (1996).
32. Cappa, S. F., Perani, D., Schnur, T., Tettamanti, M. & Fazio, F. The effects of semantic category and knowledge type on lexical-semantic access: a PET study. *Neuroimage* **8,** 350–359 (1998).
33. Mechelli, A., Sartori, G., Orlandi, P. & Price, C. J. Semantic relevance explains category effects in medial fusiform gyri. *Neuroimage* **30,** 992–1002 (2006).
34. Noppeney, U., Josephs, O., Kiebel, S., Friston, K. J. & Price, C. J. Action selectivity in parietal and temporal cortex. *Brain Res. Cogn. Brain Res.* **25,** 641 (2005).
35. Phillips, J. A., Noppeney, U., Humphreys, G. W. & Price, C. J. Can segregation within the semantic system account for category-specific deficits? *Brain* **125,** 2067–2080 (2002).
36. Kroliczak, G. & Frey, S. H. A common network in the left cerebral hemisphere represents planning of tool use pantomimes and familiar intransitive gestures at the hand-independent level. *Cereb. Cortex* **19,** 2396–2410 (2009).
37. Grossman, M. *et al.* Category-specific semantic memory: converging evidence from bold fMRI and Alzheimer's disease. *Neuroimage* **68,** 263–274 (2013).
38. Laine, M., Rinne, J. O., Hiltunen, J., Kaasinen, V. & Sipilä, H. Different brain activation patterns during production of animals versus artefacts: a PET activation study on category-specific processing. *Cogn. Brain Res.* **13,** 95–99 (2002).
39. Boronat, C. B. *et al.* Distinctions between manipulation and function knowledge of objects: evidence from functional magnetic resonance imaging. *Cogn. Brain Res.* **23,** 361–373 (2005).
40. Gorno-Tempini, M.-L. Category differences in brain activation studies: where do they come from? *Proc. R. Soc. Lond. B* **267,** 1253–1258 (2000).
41. Gerlach, C., Law, I. & Paulson, O. B. When action turns into words. Activation of motor-based knowledge during categorization of manipulable objects. *J. Cogn. Neurosci.* **14,** 1230–1239 (2002).
42. Rogers, T. T., Hocking, J., Mechelli, A., Patterson, K. & Price, C. Fusiform activation to animals is driven by the process, not the stimulus. *J. Cogn. Neurosci.* **17,** 434–445 (2005).
43. Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J. & DeYoe, E. A. Distinct cortical pathways for processing tool versus animal sounds. *J. Neurosci.* **25,** 5148–5158 (2005).
44. Canessa, N. *et al.* The different neural correlates of action and functional knowledge in semantic memory: an fMRI study. *Cereb. Cortex* **18,** 740–751 (2008).
45. Joseph, J. E., Gathers, A. D. & Piper, G. A. Shared and dissociated cortical regions for object and letter processing. *Cogn. Brain Res.* **17,** 56–67 (2003).
46. Gerlach, C. *et al.* Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia* **40,** 1254–1267 (2002).
47. Gerlach, C., Law, I., Gade, A. & Paulson, O. B. Categorization and category effects in normal object recognition: a PET study. *Neuropsychologia* **38,** 1693–1703 (2000).
48. Devlin, J. T., Rushworth, M. F. S. & Matthews, P. M. Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia* **43,** 69–74 (2005).
49. Noppeney, U., Price, C. J., Penny, W. D. & Friston, K. J. Two distinct neural mechanisms for category-selective responses. *Cereb. Cortex* **16,** 437–445 (2006).
50. Whatmough, C., Chertkow, H., Murtha, S. & Hanratty, K. Dissociable brain regions process object meaning and object structure during picture naming. *Neuropsychologia* **40,** 174–186 (2002).
51. Grafton, S. T., Fadiga, L., Arbib, M. A. & Rizzolatti, G. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage* **6,** 231–236 (1997).
52. 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).
53. Chao, L. L. & Martin, A. Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* **12,** 478–484 (2000).
54. Goldberg, R. F., Perfetti, C. A. & Schneider, W. Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* **26,** 4917–4921 (2006).
55. Bai, H. M. *et al.* Functional MRI mapping of category-specific sites associated with naming of famous faces, animals and man-made objects. *Neurosci. Bull.* **27,** 307–318 (2011).
56. Folstein, J. R., Palmeri, T. J. & Gauthier, I. Category learning increases discriminability of relevant object dimensions in visual cortex. *Cereb. Cortex* **23,** 814–823 (2013).
57. Martin, A, Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. Neural correlates of category-specific knowledge. *Nature* **379,** 649–652 (1996).
58. Perani, D. *et al.* Word and picture matching: a PET study of semantic category effects. *Neuropsychologia* **37,** 293–306 (1999).
59. Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S. & Gazzaniga, M. S. Graspable objects grab attention when the potential for action is recognized. *Nat. Neurosci.* **6,** 421–427 (2003).
60. Gerlach, C., Law, I. & Paulson, O. B. Structural similarity and category-specificity: a refined account. *Neuropsychologia* **42,** 1543–1553 (2004).
61. Wadsworth, H. M. & Kana, R. K. Brain mechanisms of perceiving tools and imagining tool use acts: a functional MRI study. *Neuropsychologia* **49,** 1863–1869 (2011).
62. Okada, T. *et al.* Naming of animals and tools: a functional magnetic resonance imaging study of categorical differences in the human brain areas commonly used for naming visually presented objects. *Neurosci. Lett.* **296,** 33 (2000).
63. 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).
64. Creem-Regehr, S. H. & Lee, J. N. Neural representations of graspable objects: are tools special? *Cogn. Brain Res.* **22,** 457–469 (2005).
65. Mruczek, R. E. B., von Loga, I. S. & Kastner, S. The representation of tool and non-tool object information in the human intraparietal sulcus. *J. Neurophysiol.* **109,** 2883–2896 (2013).
66. Zannino, G. D. *et al.* Visual and semantic processing of living things and artifacts: an fMRI study. *J. Cogn. Neurosci.* **22,** 554–570 (2010).
67. Chao, L. L., Weisberg, J. & Martin, A. Experience-dependent modulation of category-related cortical activity. *Cereb. Cortex* **12,** 545–551 (2002).
68. Anzellotti, S., Mahon, B. Z., Schwarzbach, J. & Caramazza, A. Differential activity for animals and manipulable objects in the anterior temporal lobes. *J. Cogn. Neurosci.* **23,** 2059–2067 (2011).
69. Moore, C. J. & Price, C. J. A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* **122,** 943–962 (1999).
70. Tranel, D., Martin, C., Damasio, H., Grabowski, T. J. & Hichwa, R. Effects of noun–verb homonymy on the neural correlates of naming concrete entities and actions. *Brain Lang.* **92,** 288–299 (2005).
71. Grabowski, T. J., Damasio, H. & Damasio, A. R. Premotor and prefrontal correlates of category-related lexical retrieval. *Neuroimage* **7,** 232–243 (1998).
72. Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F. & Fox, P. T. Activation likelihood estimation meta-analysis revisited. *Neuroimage* **59,** 2349–2361 (2012).
73. Martin, A. The representation of object concepts in the brain. *Annu. Rev. Psychol.* **58,** 25–45 (2007).

74. Ueno, T., Saito, S., Rogers, T. T. & Lambon Ralph, M. A. Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* **72,** 385–96 (2011).
75. 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).
76. Adlam, A. L. R. *et al.* Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? *Brain* **129,** 3066–3080 (2006).
77. Shimotake, A. *et al.* Direct exploration of the role of the ventral anterior temporal lobe in semantic memory: cortical stimulation and local field potential evidence from subdural grid electrodes. *Cereb. Cortex* **25,** 3802–3817 (2015).
78. Lambon Ralph, M. A., Lowe, C. & Rogers, T. T. Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain* **130,** 1127–1137 (2007).
79. Binkofski, F. & Buxbaum, L. J. Two action systems in the human brain. *Brain Lang.* **127,** 222–229 (2013).
80. Humphreys, G. W. & Riddoch, M. J. Features, objects, action: the cognitive neuropsychology of visual object processing, 1984–2004. *Cogn. Neuropsychol.* **23,** 156–183 (2006).
81. Humphreys, G. F. & Lambon Ralph, M. A. Fusion and fission of cognitive functions in the human parietal cortex. *Cereb. Cortex* **25,** 3547–3560 (2015).
82. Jung, J., Cloutman, L. L., Binney, R. J. & Ralph, M. A. L. The structural connectivity of higher order association cortices reflects human functional brain networks. *Cortex* http://dx.doi.org/10.1016/j.cortex.2016.08.011 (2016).
83. Embleton, K. V., Haroon, H. A., Morris, D. M., Lambon Ralph, M. A. & Parker, G. J. M. Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Hum. Brain Mapp.* **31,** 1570–1587 (2010).
84. 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).
85. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17,** 26–49 (2013).
86. Binney, R. J., Parker, G. J. M. & Lambon Ralph, M. A. Convergent connectivity and graded specialization in the rostral human temporal lobe as revealed by diffusion-weighted imaging probabilistic tractography. *J. Cogn. Neurosci.* **24,** 1998–2014 (2012).
87. Schmahmann, J. D. & Pandya, D. *Fiber Pathways of the Brain* (Oxford Univ. Press, 2009).
88. Bajada, C. J., Lambon Ralph, M. A. & Cloutman, L. L. Transport for language south of the Sylvian fissure: the routes and history of the main tracts and stations in the ventral language network. *Cortex* **69,** 141–151 (2015).
89. Bedny, M., Caramazza, A., Pascual-Leone, A. & Saxe, R. Typical neural representations of action verbs develop without vision. *Cereb. Cortex* **22,** 286–293 (2012).
90. Warrington, E. K. & McCarthy, R. Category specific access dysphasia. *Brain* **106,** 859–878 (1983).
91. Gotts, S. & Plaut, D. C. The impact of synaptic depression following brain damage: a connectionist account of 'access/refractory' and 'degraded-store' semantic impairments. *Cogn. Affect. Behav. Neurosci.* **2,** 187–213 (2002).
92. Noppeney, U. *et al.* Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain* **130,** 1138–1147 (2007).
93. Campanella, F., D'Agostini, S., Skrap, M. & Shallice, T. Naming manipulable objects: anatomy of a category specific effect in left temporal tumours. *Neuropsychologia* **48,** 1583–1597 (2010).
94. Roberts, D. *Exploring the Link Between Visual Impairment and Pure Alexia.* PhD thesis, Univ. Manchester (2009).
95. Humphreys, G. W. & Forde, E. M. E. Category specificity in mind and brain? *Behav. Brain Sci.* **24,** 497–509 (2001).
96. Laiacona, M., Capitani, E. & Barbarotto, R. Semantic category dissociations: a longitudinal study of two cases. *Cortex* **33,** 441–461 (1997).
97. Pietrini, V. *et al.* Recovery from herpes simplex encephalitis: selective impairment of specific semantic categories with neuroradiological correlation. *J. Neurol. Neurosurg. Psychiatry* **51,** 1284–1293 (1988).
98. Damasio, H., Tranel, D., Grabowski, T., Adolphs, R. & Damasio, A. Neural systems behind word and concept retrieval. *Cognition* **92,** 179–229 (2004).
99. Plaut, D. C. & Behrmann, M. Complementary neural representations for faces and words: a computational exploration. *Cogn. Neuropsychol.* **28,** 251–275 (2011).
100. Farah, M. J. & McClelland, J. L. A computational model of semantic memory impairment: modality-specificity and emergent category-specificity. *J. Exp. Psychol. Gen.* **120,** 339–357 (1991).
101. Badre, D. & Wagner, A. Semantic retrieval, mnemonic control, and prefrontal cortex. *Behav. Cogn. Neurosci. Rev.* **1,** 206–218 (2002).
102. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).
103. McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* Vol. 1 (eds Rumelhart, D. E., McClelland, J. L. & the PDP Research Group) 3–44 (MIT Press, 1986).
104. Price, C. J., Devlin, J. T., Moore, C. J., Morton, C. & Laird, A. R. Meta-analyses of object naming: effect of baseline. *Hum. Brain Mapp.* **25,** 70–82 (2005).
105. Turkeltaub, P. E. *et al.* Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp.* **33,** 1–13 (2012).
106. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9,** 97–113 (1971).
107. Eickhoff, S. B. *et al.* Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* **30,** 2907–2926 (2009).
108. Cloutman, L. L., Binney, R. J., Drakesmith, M., Parker, G. J. M. & Lambon Ralph, M. A. The variation of function across the human insula mirrors its patterns of structural connectivity: evidence from *in vivo* probabilistic tractography. *Neuroimage* **59,** 3514–3521 (2012).
109. Visser, M. & Lambon Ralph, M. A. Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *J. Cogn. Neurosci.* **23,** 3121–3131 (2011).
110. 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).
111. Parker, G. J. M. & Alexander, D. C. Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Phil. Trans. R. Soc. B* **360,** 893–902 (2005).
112. Rohde, D. L. T. *LENS: The Light, Efficient Network Simulator*. Technical Report CMU-CS-99-164 (Carnegie Mellon Univ., 1999).
113. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Learning Representations by Back-Propagating Errors* (MIT Press, 1988).
114. Garrard, P. & Carroll, E. Lost in semantic space: a multi-modal, non-verbal assessment of feature knowledge in semantic dementia. *Brain* **129,** 1152–1163 (2006).
115. Dixon, M. J., Bub, D. N. & Arguin, M. The interaction of object form and object meaning in the identification performance of a patient with category-specific visual agnosia. *Cogn. Neuropsychol.* **14,** 1085–1130 (1997).
116. Dixon, M. J., Bub, D. N., Chertkow, H. & Arguin, M. Object identification deficits in dementia of the Alzheimer type: combined effects of semantic and visual proximity. *J. Int. Neuropsychol. Soc.* **5,** 330–345 (1999).
117. Zevin, J. D. & Seidenberg, M. S. Simulating consistency effects and individual differences in nonword naming: a comparison of current models. *J. Mem. Lang.* **54,** 145–160 (2006).

Acknowledgements

This research was supported by a programme grant from the Medical Research Council (MRC, UK, MR/J004146/1) to M.A.L.R. and by a University Fellowship from UW-Madison to L.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank L. Cloutman for assisting with the tractography analysis and R. Ishibashi for assisting with the ALE analysis.

Author contributions

All authors contributed to the entire process of this project, including project planning, experiment work, data analysis and writing the paper.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.C., M.A.L.R. or T.T.R.

How to cite this article: 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).

Competing interests

The authors declare no competing interests.