A unified model of human semantic knowledge and its disorders

DOI:
10.1038/s41562-016-0039

Document Version
Accepted author manuscript

Title: A unified model of human semantic knowledge and its disorders

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

Affiliations:
[1]Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI 53705, USA

[2]Neuroscience and Aphasia Research Unit (NARU), Division of Neuroscience & Experimental Psychology, School of Biological Sciences, University of Manchester, M13 9PL, UK

[3]Stanford Cognitive and Systems Neuroscience Laboratory, 1070 Arastradero Rd. Suite 220, Palo Alto, CA 94304, USA

Summary: 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.

Text:

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: (a) the system is modular and domain-specific, with components that have evolved to support different knowledge domains[2,3], e.g. animals, tools, people, etc., or (b) it is interactive and domain-general, with all components contributing to all knowledge domains[4–6]. Despite profoundly different implications about 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* - 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 neuro-

cognitive 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[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. The current paper tests 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 behavior 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 (a) patterns of functional activation in brain imaging studies, (b) impairments observed in the primary disorders of semantic representation, and (c) the anatomical bases of these disorders.

Activation likelihood estimate (ALE) analysis

Prior empirical, modeling, and neuroimaging work (SI-Discussion 1) has identified several cortical regions that contribute to semantic processing and their respective functional roles, including: (1) the posterior fusiform gyrus (pFG), which encodes visual representations of objects[18,19]; (2) the superior temporal gyrus (STG), which encodes auditory representations of speech[20]; (3) lateral parietal cortex, which encodes representations of object function and praxis[19,21,22]; and (4) 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. Like 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 to date (details in 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 Table S1). Results are shown in Figure 1 and Figure S1.

-- Figure 1 about here --

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, though the animal vs. artifact contrast was only significant in the right, possibly because the homologous left-hemisphere region is "sandwiched" between two areas showing the reverse pattern (pMTG and medial pFG; see Fig. S2). 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, though 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[78] have established the importance of ventral ATL for domain-general semantic processing. Prior models[5,12,79] included ventral ATL as a cross-modal semantic hub (see SI-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, consistent with the proposal that this region encodes representations of object-directed action[19,80]. One prior model incorporated function representations in the lateral parietal cortex[12]. The cluster spanned inferior and superior parietal lobes (IPL and SPL), which patient and imaging literatures suggest encode different aspects of action knowledge[80]. Thus we included both as separate regions of interest in the connectivity analysis and the model.

Posterior middle temporal gyrus (pMTG) exhibited more activations for artifacts than 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.

Lateral occipital complex (LOC) activated more often for animals than artifacts, which probably reflects domain differences in visual structure including greater complexity and more overlap among animals relative to manmade objects[81]. We thus identify LOC as a source of visual input to 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[82,83], 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 entral ATL[84]. Seeds were placed in the white matter underlying the regions of interest from the ALE analysis or the literature (Fig. 1 and 2; for ROI definition, see Methods), mapped back to native space for each subject[85]. STG and LOC were excluded from the analysis since their connectivity is well-studied[86,87] and they are posited to

provide spoken-word and visual input, respectively, to the semantic network. Results are shown in Figure 2.

-- Figure 2 about here --

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 Table S2; for thresholding, see Methods). ATL also projected to both pFG regions and to the pMTG (>2.5% in more than half 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 participants).

Temporo-parietal connections. Streamlines from the ATL did not extend into parietal cortex as also found previously[83,88]. Streamlines from pMTG, however, projected both to ATL and to IPL, providing an indirect route from IPL to the 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. Recent neuroanatomical studies from MR 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[89,90]—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 neighborhood 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 3 about here --

Figure 3A shows a schematic of the ALE and connectivity results. We next constructed a neurocomputational model whose architecture mirrors these results, shown in Figure 3B. The model is a deep recurrent neural network that computes mappings amongst visual representations of objects (coded in LOC), verbal descriptors (STG), and functional (IPL) and praxic (SPL) action representations[85]. The model was trained with predictive error-driven learning to generate an item's full complement of visual, verbal, function and praxic properties, given a subset of these as input. Surface representations were generated to capture three well-documented aspects of environmental structure: (a) hierarchical similarity with few properties shared across domains, more shared within domains, and many shared within basic categories[11]; (b) many more praxic and functional features for artifacts and somewhat more visual features for animals[10,11]; and (c) more feature overlap amongst animals than artifacts[5] (see SI-Methods 5 and Table S3)

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 analogs of fifteen 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 analogs of the BOLD response and analyzed to identify model regions showing systematic category effects[12] (see Methods).

Results are shown in Figure 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 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[9,63,91]. 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[9,63,91].

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 (a) reflect degraded semantic representation rather than access/retrieval deficits[92,93] and (b) have been shown in case-series studies to manifest predictable patterns of impairment. These include: (a) semantic dementia (SD), where progressive bilateral ATL atrophy produces a category-general semantic impairment[79]; (b) herpes simplex viral encephalitis (HSVE), where acute bilateral ATL pathology produces chronic impairments disproportionately affecting animals[94]; (c) temporo-parietal tumor resection (TPT), which produces greater impairment for artifacts[95]; and (d) forms of visual agnosia (VA) producing slower and less accurate recognition for animals[81,96].

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 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 vs. HSVE[94] (see SI-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[85].

-- Figure 4 about here --

Results are shown in Figure 4. The model captures the direction and magnitude of several key phenomena including: (a) no category effect in SD, (b) a substantial animal disadvantage in both HSVE variants (results of HSVE+ in Fig. S5), (c) a modest artifact disadvantage in TPT, (d) an animal disadvantage in response time in VA, (e) worse anomia in SD than HSVE, and (f) a smaller and opposite category effect in TPT compared to 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"[81]: 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[97]. In HSVE, the model pathology is identical to SD—the category effect thus arises through re-learning, via two mechanisms. First, intact functional and praxic layers can support new learning for items with these properties, that is, for artifacts. Second, because animals share more properties than artifacts their ATL representations are also "conceptually crowded," compromising relearning of inter-item differences when ATL representations are damaged[79]. 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 Figs. S4 and S5).

The simulation suggests a novel resolution to the long-standing puzzle of why patients with HSVE and SD show qualitatively distinct impairments despite largely overlapping pathology[79]. 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 SI-Discussion 3.2 and 4). To assess this hypothesis, we first evaluated model predictions by regressing the magnitude of the category effect (artifact accuracy – 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 SI-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 shrunk the category effect (Fig. S7A&B; interaction for HSVE $t = 2.501$, $p = .014$; interaction term for HSVE+ $t = 2.137$, $p = .035$). We then assessed whether the same pattern is observed in the literature. Across 19 previously-reported HSVE cases of category-specific impairment (Table S6), 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 [$t = 3.298$, $p < .01$]: relearning produced larger effects when deficits were severe but smaller effects when deficits were mild (Fig. S7C). The same pattern was also observed longitudinally in 4 patients with HSVE[98,99] (Fig. S7D). By contrast, this pattern was not found in the non-HSVE cases (for full results, see Table S7). 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 while artifact-selective deficits occur with temporo-parietal pathology[100]. 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 vs 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, while 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[19,95,101], 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 (a) category-specific patterns of functional activation in sighted and congenitally-blind individuals, (b) patterns of impairment observed across four different neuropsychological syndromes, and (c) 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[10,102]; 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 facilitate 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. Like the correlated-structure view, category-selectivity arises partly from different patterns of overlap among animal versus artifact properties[6], but in our model network 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 SI-Discussion 2 for relationship to other models).

In emphasizing semantic representation we have not considered the fronto-parietal systems involved in semantic control[103], 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 crucial 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 and healthy and disordered cognitive functioning[74]. While interest in neural networks has recently rekindled in machine learning[104], their original promise as tools for bridging minds and brains[105] 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 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 [25,106] and found 49 papers describing 73 independent studies (31 for animal and 42 for artifact; for study selection, see SI-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[107,108]. The ALE analysis strictly followed the steps proposed by Price et al.[106] and Eickhoff et al.[72,107,108], and coordinates in MNI space were used for ALE analysis and reports. Main effects of animal and artefact concepts (concordance of foci showing greater activations for animal vs. baseline and artefact vs. baseline) are reported in Table S1. 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 University of Manchester, UK[88]. All participants are right-handed as determined by the Edinburgh Handness Inventory[109]. Inclusion and exclusion criteria were stated in previous studies[88,110], 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 8 element SENSE head coil. Diffusion weighted imaging was performed using a pulsed gradient spin echo echo-planar sequence with TE=59 ms, TR≈1500 ms (cardiac gated), G=62 mTm−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=1200 smm−2 (Δ=29.8ms, δ=13.1ms), 1 at b=0, and SENSE acceleration factor=2.5. A high-resolution T1-weighted 3D turbo field echo inversion recovery scan (TR≈2000 ms, TE=3.9 ms, TI=1150ms, 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[88,110].

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 an fMRI study[111] that reported cross-modal activation for conceptual processing in the ATL. Due 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 to this region slowed naming of tools but not animals[112]. 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 anterior-most 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 SI-Methods 2.

Probabilistic tracking procedure. We restricted our analysis to the left hemisphere, and following similar procedure of previous study[110], a sphere with a diameter of 6mm centered on the seed coordinate for each ROI was then drawn in the MNI template (see Table S2 for the exact coordinates; details in SI). 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 PICo (Probabilistic Index of Connectivity) method[110,113]. Step size was set to 0.50 mm. Stopping criteria for the streamlines were set so that tracking terminated if 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[110]. 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 in SI-Methods 3). A group-averaged tractography image was then obtained by averaging the normalized individual data[110].

Computer simulations of fMRI data. The model architecture shown in Figure 3B (main text) was implemented using the Light Efficient Network Simulator (LENS) software[114]. The model included four visible layers directly encoding model analogs of visual, verbal (names and descriptions), praxic, and functional properties of objects. Each visible layer was reciprocally connected with its own modality-specific hidden layer, providing model analogs to the posterior fusiform (pFG, visual hidden units), superior temporal gyrus (STG, verbal hidden units), inferior parietal lobule (IPL, function hidden units), and superior parietal lobule (SPL, praxic 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 Figure 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] (details see SI-Methods 4), to capture the observation that medial pFG is more strongly connected to parietal regions than is lateral pFG[19]. All units employed 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 praxic representations for 24 different exemplars of animals and 24 different exemplars of tools, with each domain organized into 4 basic categories, each containing 6 exemplars (for representational schemes of training exemplars, see Table S3 and SI-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 (a) items in different domains shared few properties; (b) items within the same category shared many properties; (c) animals from different categories shared more properties than did artifacts from different categories; and (d) animals had more properties overall than did artifacts. Each item was also given a unique name as a 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 praxic 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 analog of vision-to-action affordances. Function representations simply duplicated the praxic 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 praxic features, similar to our previous work[12] (details see SI-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[115]. '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 (details, see SI-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 Table S4).

Simulating functional brain imaging studies. The brain imaging studies simulated involved two tasks: picture viewing, in which participants made a semantic judgment from a picture of a familiar item, and name comprehension, in which they made a semantic judgment 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 analog 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 vs. 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 Artifacts concepts using the averaged data across the 15 model subjects. An ROI analysis was then performed on activation value of the peak unit averaged together with two neighboring units on either side.

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 in SI-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 Fig. S3).

(1) Semantic dementia (SD) is a neurodegenerative disorder associated with gradual thinning of cortical grey matter and associated white-matter fibers, centered in the ATL[78], and produces a robust, progressive and yet selective deterioration of semantic knowledge for all kinds of concepts, across all modalities of reception and expression[76,5,116]. 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/reorganize.

(2) Herpes Simplex Viral Encephalitis (HSVE) is a disease that produces rapid bilateral necrosis of gray and white matter, generally encompassing the same regions affected in SD, but patients with semantic impairments from HSVE, has been found with greater damage in temporal white matter especially in the lateral axis[94]. HSVE patients often show less semantic impairment overall, with greater deficits of knowledge for animals than for manmade objects[79,94]. The main paper considers two potential explanations, each associated with a different model of HSVE pathology. The model first captures differences in the time-course of SD vs. HSVE: whereas the former progresses slowly over a course of years, the latter develops rapidly and is then halted by anti-viral 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 SI-Discussion 3.2). Retraining employed 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 Fig. S4 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[94]. 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 Fig. S5). 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 Fig. S6).

(3) Temporo-parietal tumor resection (TPT). Campanella and colleagues[95] 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 tumors. The group exhibited significantly worse knowledge of nonliving things compared to animals, with difference scores in naming accuracy ranging from 2%-21%. Voxel-based lesion-symptom mapping (VLSM) 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 visual agnosia (VA). Finally, a long tradition of research suggests that forms of associative visual agnosia arising from damage to occipitotemporal regions can have a greater impact on recognition of living than nonliving things[117,118]. The deficit is specific to vision, and more evident in naming latency rather than accuracy at milder impairment[96]. 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 in order 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 (a) the most active of all basic name units and (b) was activated above 0.5; otherwise it was scored as incorrect. In visual agnosia 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 was the number of ticks for naming without any lesion in the model[119]. 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 animal and artifact categories. See Table S5 for naming accuracy and latency at different levels of lesion severity measured as percentage of affected connections.

Data availability. Program scripts and source data that support the data analysis of this project are available in online public repositories, and more details are available upon request. See https://github.com/halleycl/ChenETAL_NatHumanBehav_SI-Online-materials and https://app.box.com/v/ChenETAL-NatHumanBehav-SI.

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,** 964–8 (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. U. S. A.* **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 (80-.).* **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. London. Ser. B Biol. Sci.* **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,** 434–45 (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 (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,** 2883–96 (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,** 3802–17 (2015).
78. Acosta-Cabronero, J. *et al.* Atrophy, hypometabolism and white matter abnormalities in semantic dementia tell a coherent story. *Brain* **134,** 2025–2035 (2011).
79. 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).
80. Binkofski, F. & Buxbaum, L. J. Two action systems in the human brain. *Brain Lang.* **127,** 222–229 (2013).
81. Humphreys, G. W. & Riddoch, M. J. Features, objects, action: The cognitive neuropsychology of visual object processing, 1984–2004. *Cogn. Neuropsychol.* **23,** 156–183 (2006).
82. Humphreys, G. F. & Lambon Ralph, M. A. Fusion and Fission of Cognitive Functions in the Human Parietal Cortex. *Cereb. Cortex* **25,** 3547–3560 (2015).
83. 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* (2016).
84. 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).
85. Materials and methods are available as supplementary materials on Science Online.
86. 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,** 2728–38 (2010).
87. 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).
88. 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).
89. Schmahmann, J. D. & Pandya, D. *Fiber pathways of the brain*. (Oxford University Press, 2009).
90. 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,** 141–51 (2015).
91. Bedny, M., Caramazza, A., Pascual-Leone, A. & Saxe, R. Typical neural representations of action verbs develop without vision. *Cereb. Cortex* **22,** 286–93 (2012).
92. Warrington, E. K. & McCarthy, R. Category specific access dysphasia. *Brain* **106,** 859–878 (1983).
93. Gotts, S. & Plaut, D. C. The impact of synaptic depression following brain damage: A connectionist account of ``access/refractory'' and ``degraded-store'' semantic impairments. *CABN* **2,** 187–213 (2002).
94. Noppeney, U. *et al.* Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain A J. Neurol.* **130,** 1138–1147 (2007).
95. Campanella, F., D'Agostini, S., Skrap, M. & Shallice, T. Naming manipulable objects: anatomy of a category specific effect in left temporal tumours. *Neuropsychologia* **48,**

1583–97 (2010).
96. Roberts, D. Exploring the link between visual impairment and pure alexia. (University of Manchester, 2009).
97. Humphreys, G. W. & Forde, E. M. E. Category specificity in mind and brain? *Behav. Brain Sci.* **24,** 497–509 (2001).
98. Laiacona, M., Capitani, E. & Barbarotto, R. Semantic category dissociations: A longitudinal study of two cases. *Cortex* **33,** 441–461 (1997).
99. 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).
100. Damasio, H., Tranel, D., Grabowski, T., Adolphs, R. & Damasio, A. Neural systems behind word and concept retrieval. *Cognition* **92,** 179–229 (2004).
101. Plaut, D. C. & Behrmann, M. Complementary neural representations for faces and words: a computational exploration. *Cogn. Neuropsychol.* **28,** 251–275 (2011).
102. 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).
103. Badre, D. & Wagner, A. Semantic retrieval, mnemonic control, and prefrontal cortex. *Behav. Cogn. Neurosci. Rev.* **1,** 206–218 (2002).
104. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).
105. McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (eds. Rumelhart, D. E., McClelland, J. L. & the PDP Research Group) **1,** 3–44 (MIT Press, 1986).
106. 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).
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. 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).
109. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9,** 97–113 (1971).
110. 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).
111. 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).
112. 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).
113. Parker, G. J. M. & Alexander, D. C. Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos. Trans. R. Soc. B Biol. Sci.* **360,** 893–902 (2005).
114. Rohde, D. L. T. LENS: the light, efficient network simulator. *Technical Report CMU-CS-*

99-164 (1999).
115. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Learning representations by back-propagating errors*. (MIT Press, Cambridge, MA, USA, 1988).
116. 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).
117. 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).
118. 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).
119. 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).

Supplementary Information is available in the online version of the paper.

Acknowledgments

This research was supported by a programme grant from the Medical Research Council (MRC, UK, MR/J004146/1) to MALR and by a University Fellowship from UW-Madison to LC. We also want to thank Lauren Cloutman for assisting the tractography analysis and Ryo Ishibashi for assisting the ALE analysis.

Author contributions

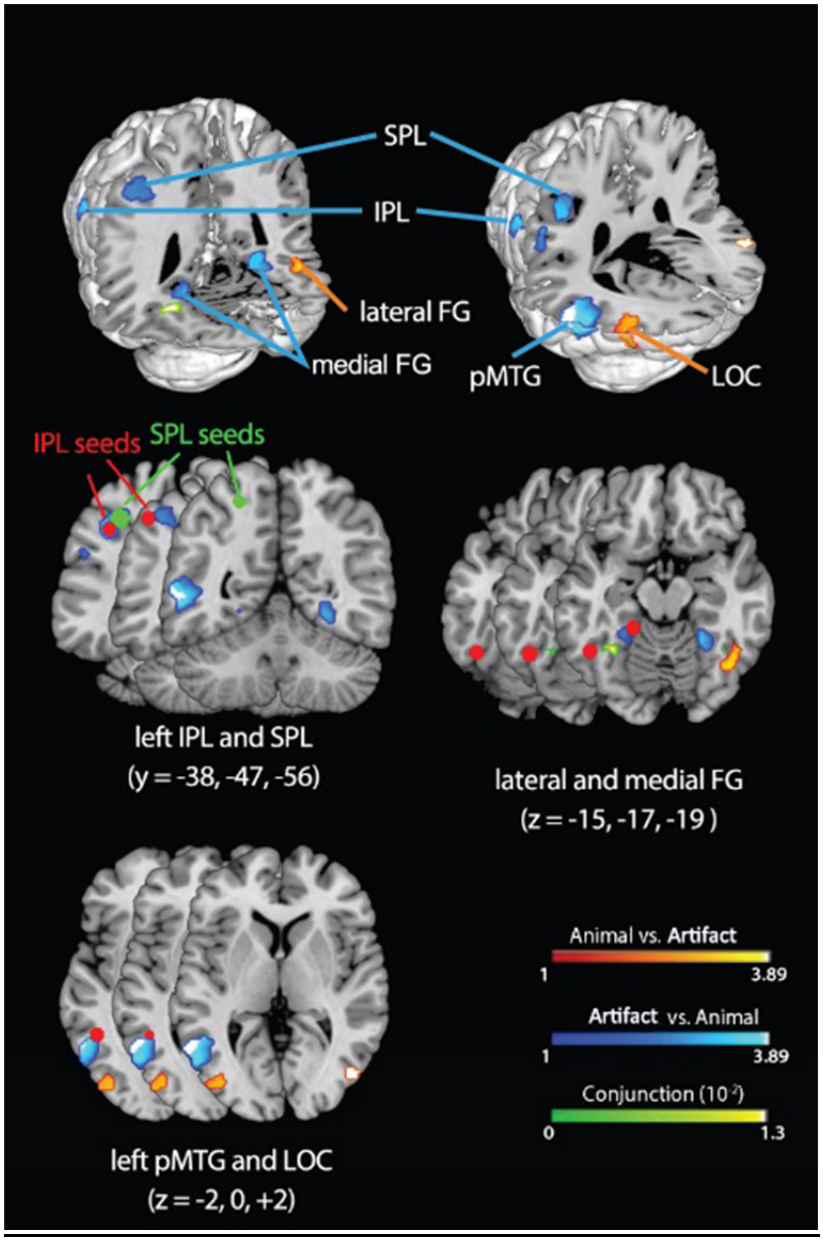
All authors contributed to the entire process of this project, including project planning, experiment work, data analysis, and writing the paper.

Competing financial interests

The authors declare no competing financial interests.

Materials and Correspondence

Correspondence and requests for materials should be addressed to Lang Chen (lchen32@stanford.edu), Matthew A. Lambon Ralph (matt.lambon-ralph@manchester.ac.uk), or Timothy T. Rogers (ttrogers@wisc.edu).



848

849 **Figure 1** ALE analysis showing regions that systematically respond more to animals than
850 artifacts (orange), more to artifacts than animals (blue), or equally to both (green). Red dots
851 indicate seed points from activation likelihood estimation (ALE) analysis and literature review.
852 IPL = inferior parietal lobe, SPL = superior parietal lobe, pFG = posterior fusiform gyrus, pMTG
853 = posterior middle temporal gyrus, LOC = lateral occipital complex.
854
855

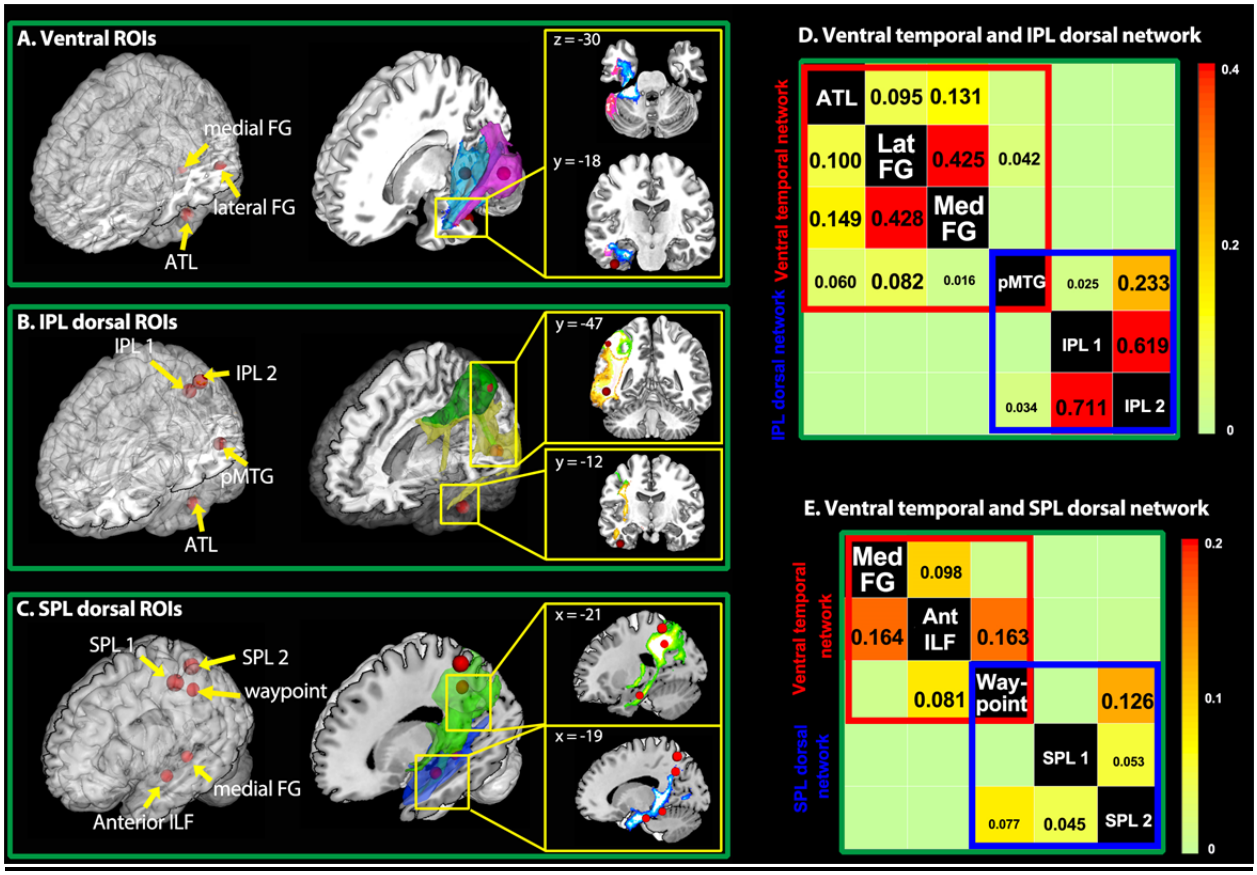


Figure 2 Tractography results. Red spheres indicate seed points from activation likelihood estimation (ALE) analysis and literature review. (**A).** Streams from medial (blue) and lateral (pink) pFG project to ATL. (**B).** Streams from pMTG (yellow) project to ATL and IPL, while IPL streams (green) project to pMTG but not ATL. (**C).** Streams from inferior ATL white matter (blue) pass by medial pFG and branch superiorly, where they intersect SPL streamlines (green). The waypoint seed was placed at this intersection. **(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.

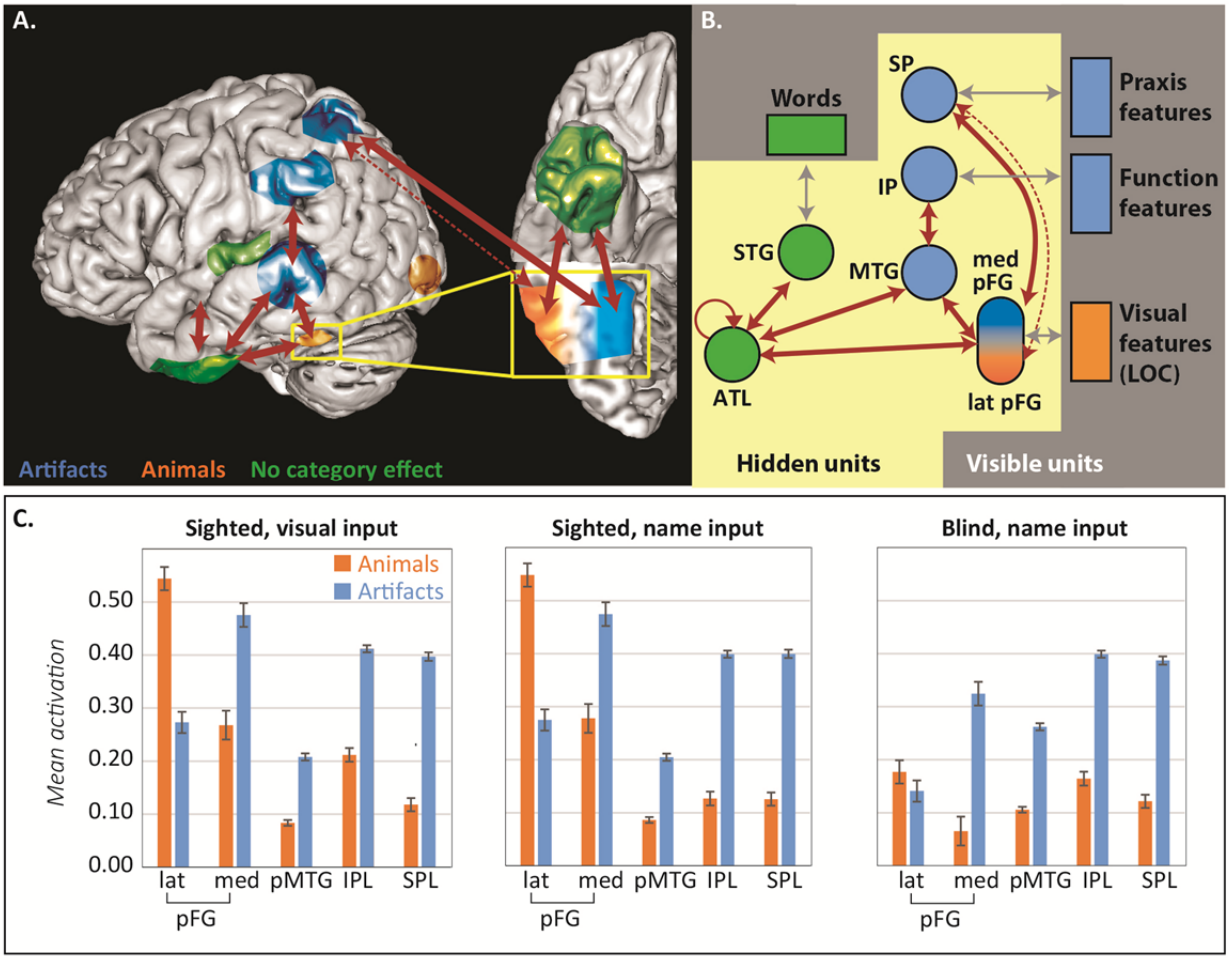


Figure 3 Model architecture and fMRI data simulations. (**A).** Schematic showing ALE and connectivity results. Red arrows indicate significant connectivity in tractography while colors indicate semantic category effects in the ALE analysis. The dotted arrow indicates that connectivity diminishes from 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 analogs of cortical regions of interest 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 color indicates expected category effects using the same scheme as panel A. Red arrows indicate model connections that correspond to tractography results; gray arrows indicate connections that mediate activation between visible and hidden units. (**C).** Mean unit activation for animals or artifacts in each model region of interest, for visual and word inputs of the "sighted" model (left and middle) and for word inputs in the "blind" model (right).

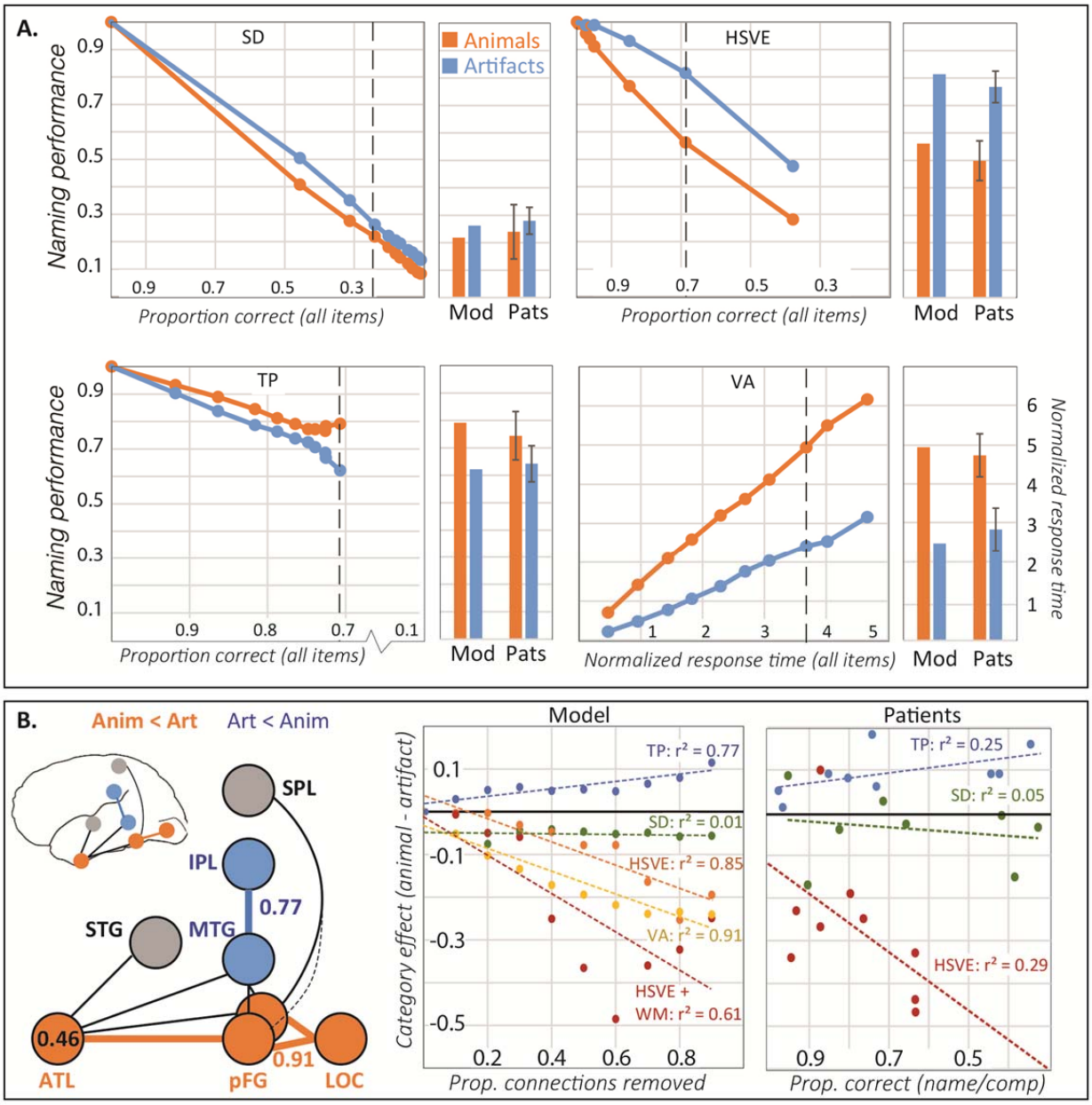
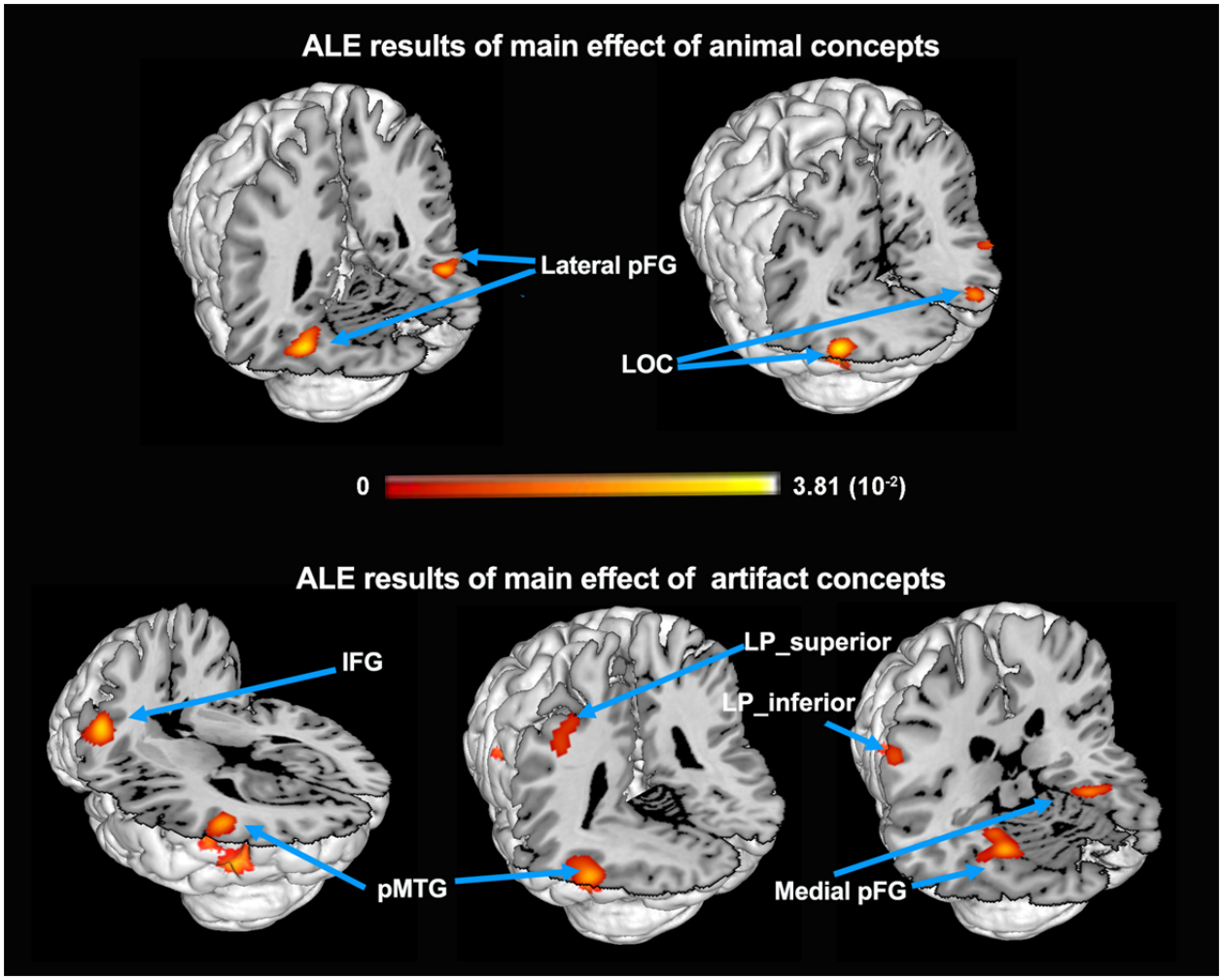


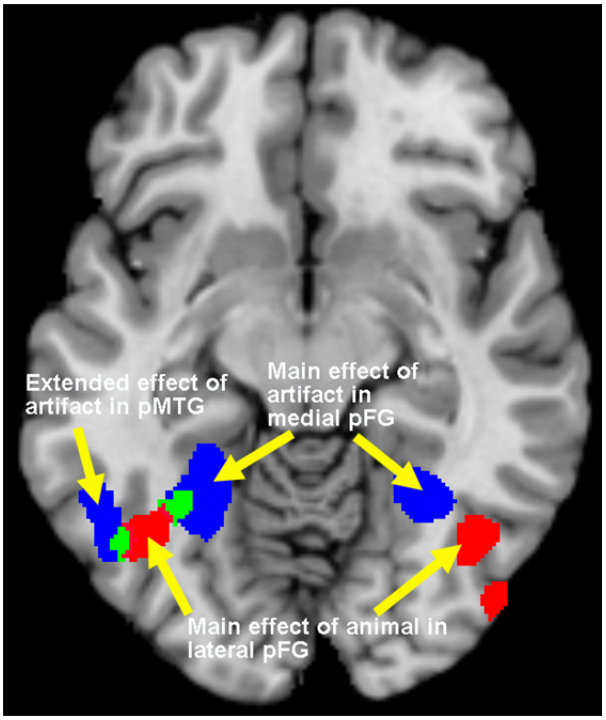
Figure 4 Results of patient simulations. (**A).** Line plots show model naming accuracy for animals and artifacts at ten increasing levels of damage for each disorder plotted against overall accuracy (all items). Dashed vertical lines indicate the damage level that most closely matches mean overall accuracy in the corresponding patient group. HSVE data are for the homogeneous damage model (HSVE); data for the asymmetric damage model (HSVE+) appear in Supplementary Figure S5. Barplots show accuracy by category for the model at this level compared to patient means/standard errors reported in [79,95,96]. (**B).** Lesion-symptom mapping results. Left: Layers/connections where lesion size predicts increasing artifact (blue) or animal (orange) disadvantage. 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.

Supplementary Information

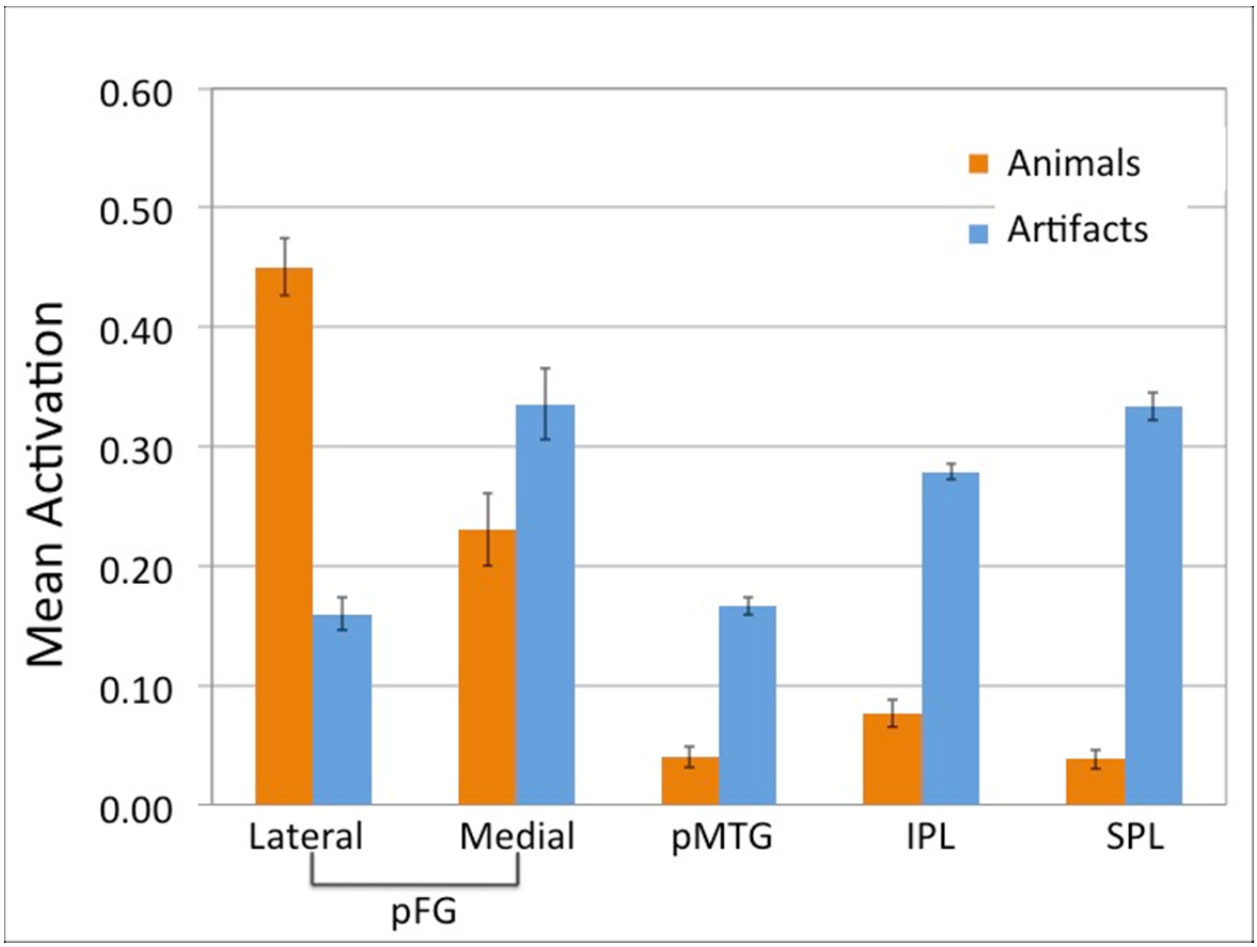
Table of Contents



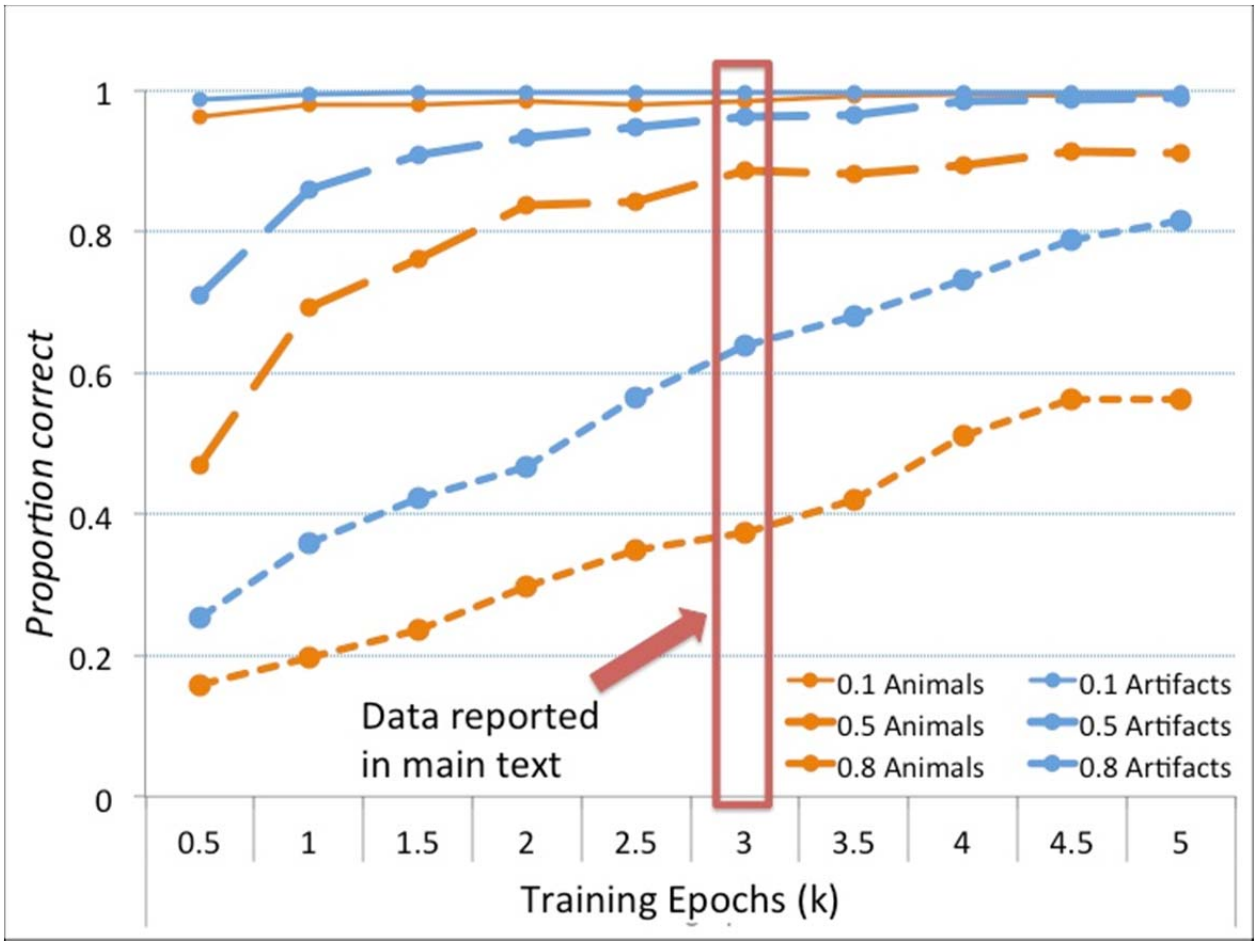
Supplementary Figure S1. Main effect of animal and artifact concepts in the ALE meta-analysis showing clusters where animal concepts (top) or artifact concepts (bottom) elicit reliably more activation than a control condition. Note that the animal effect in lateral pFG is observed in both hemispheres. pFG = posterior fusiform gyrus; LOC = lateral occipital cortex; pMTG = posterior middle temporal gyrus; IFG = inferior frontal gyrus; LP = lateral parietal cortex.



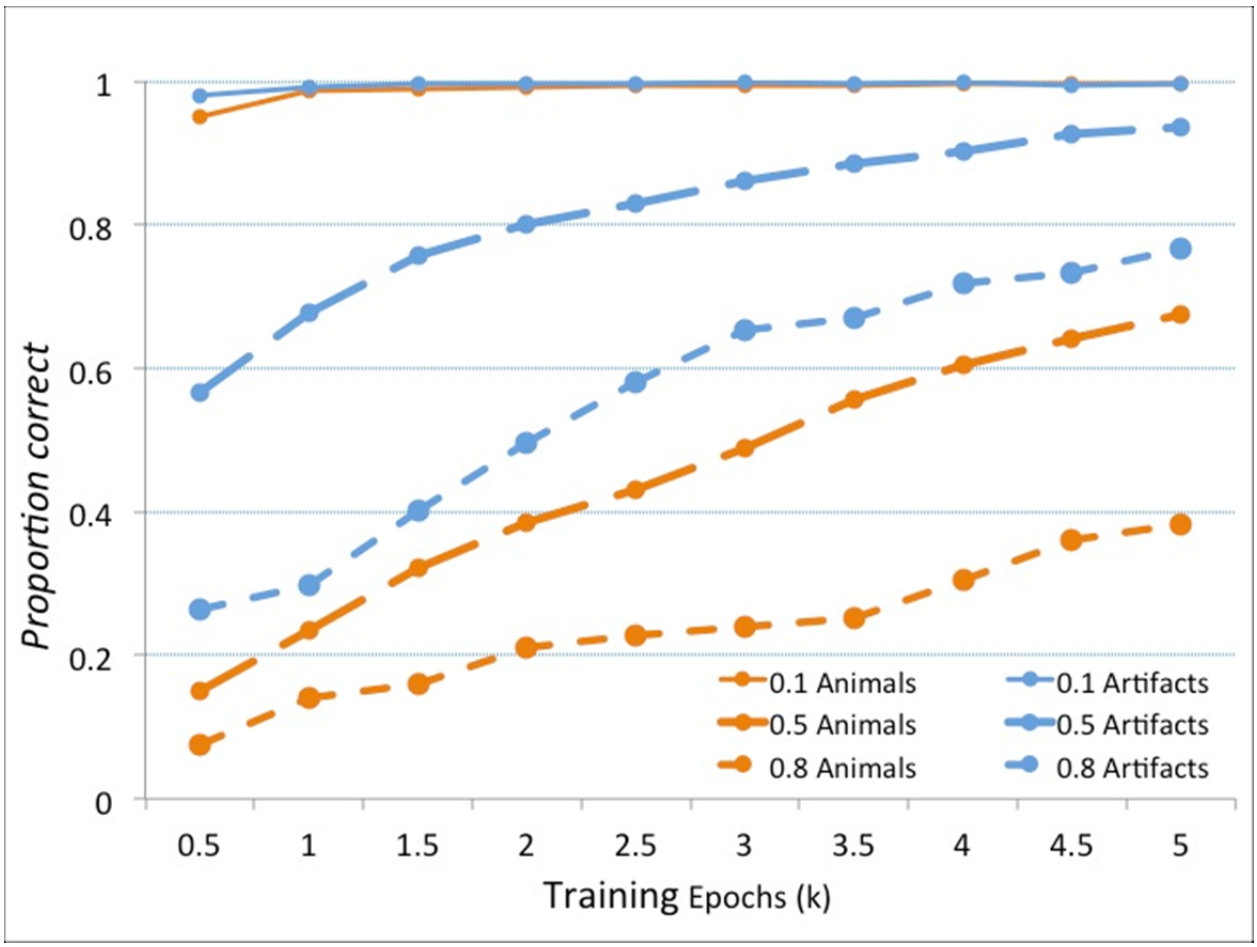
Supplementary Figure S2. Overlapping main effects of animal and artifact concepts in left pFG. The blue clusters showed a significant effect for artifacts in the ALE analysis while the red clusters show significant clusters for animal concepts. Green shows the overlap between these, also revealed in the conjunction analysis reported in the main text. Of note, animals but not artifacts activated the lateral fusiform bilaterally in this analysis, a finding consistent with several prior studies[1]. The lateral pFG contrast of animal to artifact activation was only significant in the right hemisphere. This figure suggests why: in the left hemisphere the artifact-selective activations in medial pFG and pMTG bracket the animal-selective activation in lateral pFG. Thus bleed-over from these foci may have produced an insignificant category contrast in lateral fusiform. Nevertheless, the main effect in the ALE analysis and the robust contrast effect in the right hemisphere jointly support the conventional claim that lateral pFG yields category-sensitive activation for animal concepts.



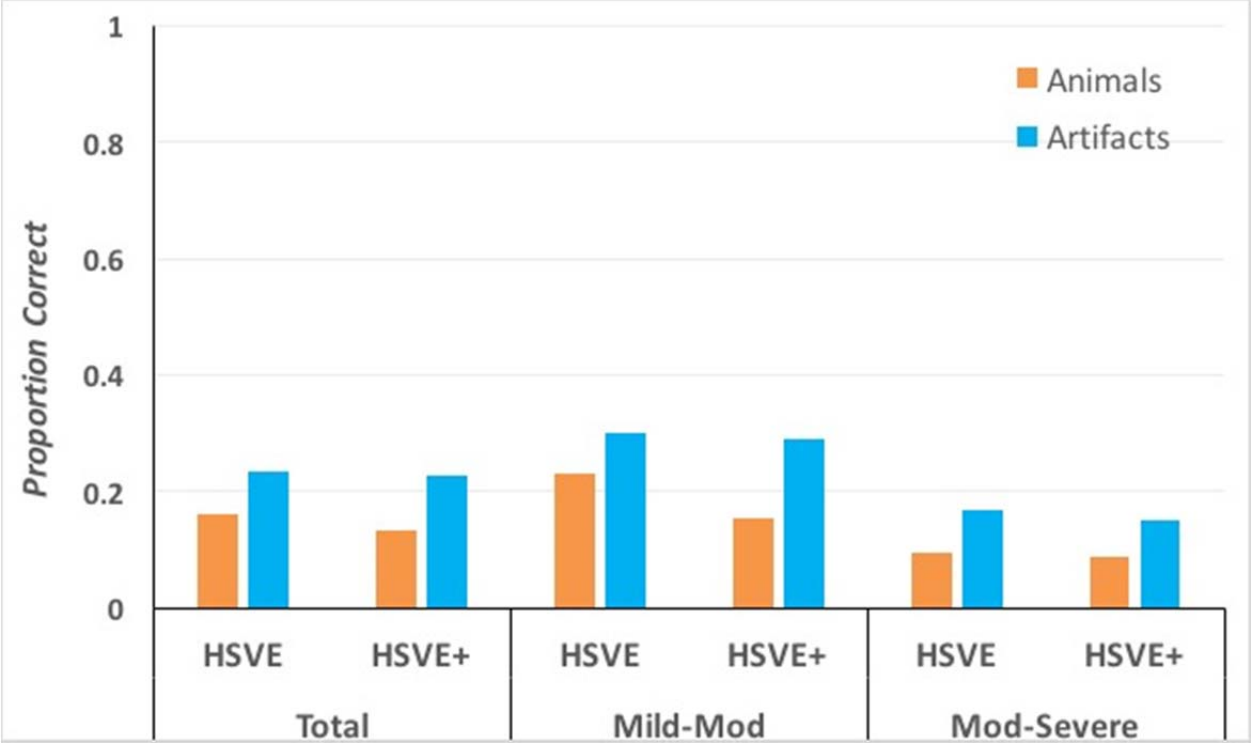
Supplementary Figure S3. Replication of fMRI data simulation for the sighted model performing the visual viewing task. In the model for simulating patient data, we balanced the exposure frequency to animal and artifact object names. Here, we demonstrate that this model simulation still accounts for the category-specific activation patterns in the ALE analysis.



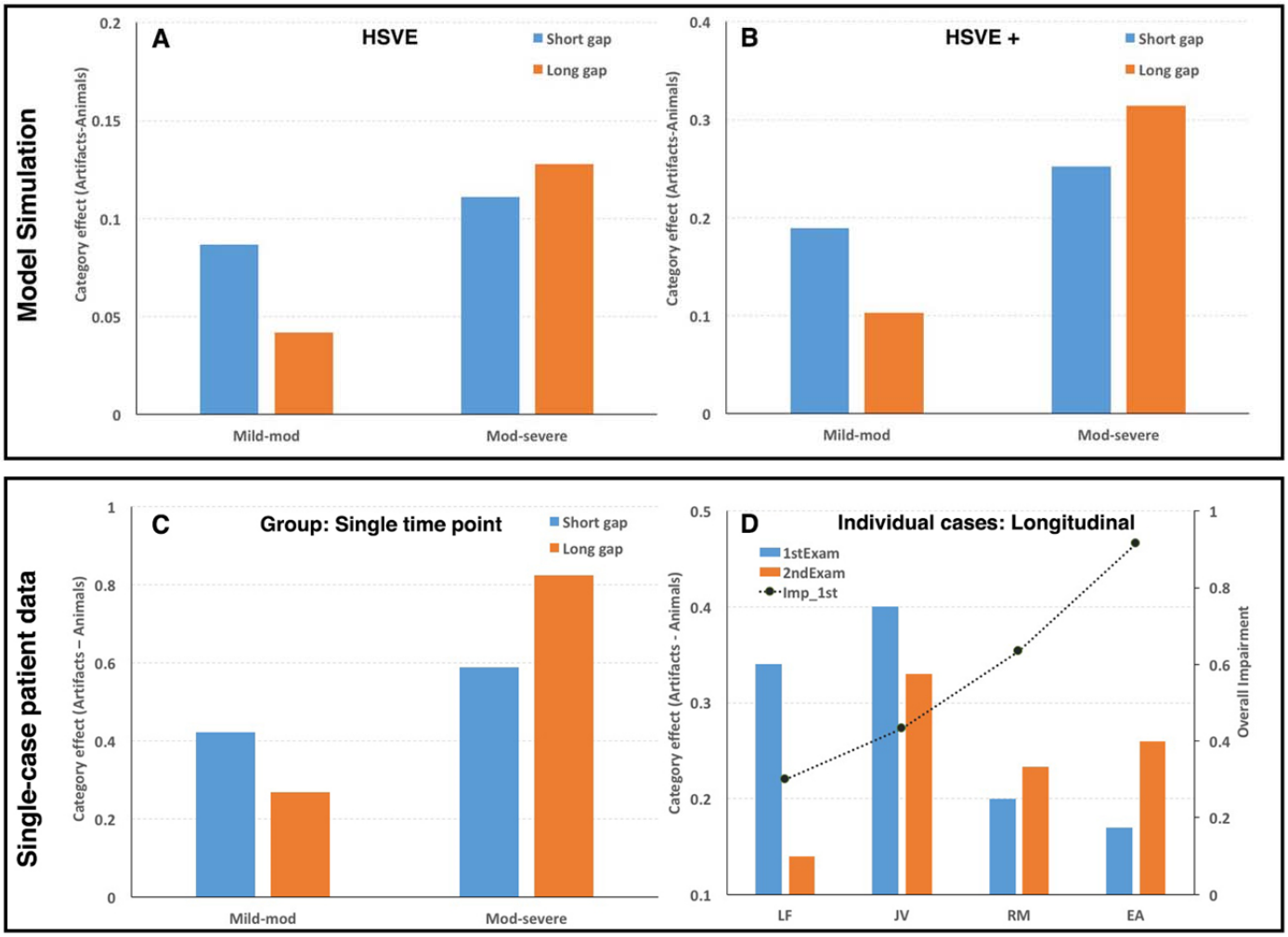
Extended Date Figure S4. Relearning trajectories for model simulations of HSVE at three representative levels of lesion severity. After the sighted models were fully trained, connections associated with ATL semantic hub were randomly pruned at three levels of severity (HSVE variant): 10% (thin solid lines), 50% (thick dashed lines) and 80% (normal segmented lines). The models relearned both domains fairly rapidly, but with a clear disadvantage for animal concepts, especially at more severe levels of initial lesion (80% affected connections). The red rectangle marked the point during recovery where the data were drawn for the presentations in the main text.



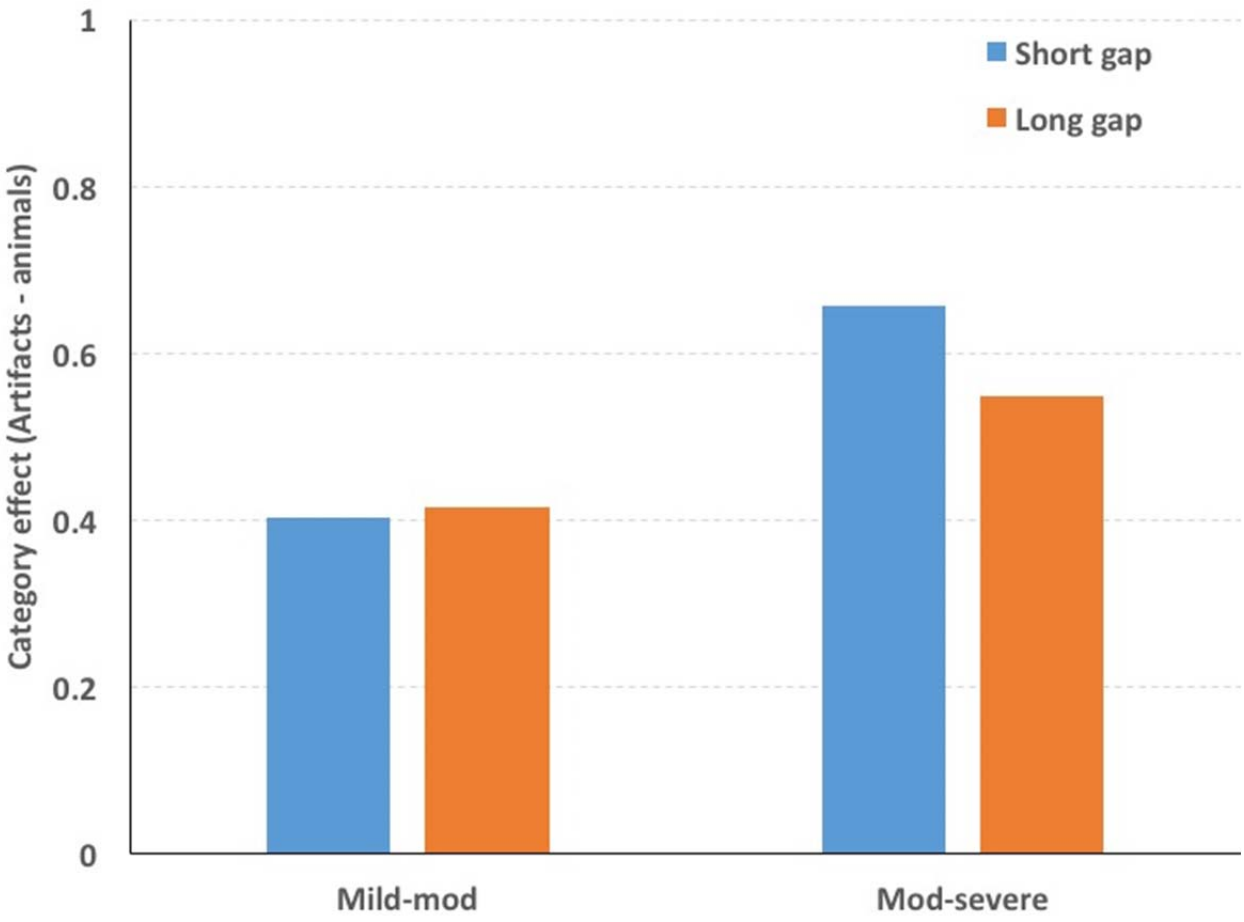
Supplementary Figure S5. Relearning trajectories of model simulation for HSVE patients with additional lesion in lateral pFG-to-ATL connections (HSVE+ variant) at three representative levels of lesion severity. Similar relearning was observed but the recovery gap between animal and manmade knowledge was larger with the increased probability of lesioning lateral pFG-to-ATL connections.



Supplementary Figure S6. Category-specific impairment in HSVE cases prior to relearning. Naming was severely impaired in both model variants. When the lesion was applied to connections associated with ATL hub homogeneously (HSVE variant), a very small disadvantage was observed for animals, but this category effect increased when damage disproportionately affected the ATL-and-lateral-FG connections (HSVE+ variant), especially for milder lesions (proportion of connections affected from 0.1 to 0.5). When the lesion was more severe (affected proportion from 0.6 to 1), both HSVE and HSVE+ variants showed a floor effect and little difference between naming animals and artifacts.



Supplementary Figure S7. Predicted and observed effects of relearning on the magnitude of the category effect in HSVE. (A) Model prediction from the HSVE variant. When the initial lesion was mild to moderate (affected connections <=0.5), the long gap/more relearning (>2.5k; orange) lead to smaller category-specific effect than short gap/less relearning (<=2.5k; blue); however, when the initial lesion was moderate to severe (affected connections >0.5), the long gap/more relearning enlarged the size of category-specific effect. (B) A similar prediction was made from the HSVE+ variant. (C) Group averaged result at initial examination of HSVE cases reviewed in Capitani et al[2]. The observed pattern is consistent with the model prediction that at mild-moderate level of overall impairment (error rate <=0.5), long gap (> 1year; orange) between the disease insult and initial examination lead to smaller category-specific effect than short gap (<= 1 year; blue); but the reverse was observed when overall impairment was moderate to severe (error rate > 0.5). (D) A similar pattern was observed in four HSVE cases who were examined more than once[3,4]. When the initial impairment was less severe (cases LF and JV), the category-specific effect reduced over time, whereas the reverse was observed for cases who had a more severe initial impairment (RM and EA). Blue bars show the category effect on 1st examination and orange bars on the 2nd. The black dashed line denotes the overall impairment at 1st examination.



Supplementary Figure S8. The different pattern of time, overall impairment and category effect in non-HSVE patients. The bars show the mean category effect across patients with semantic impairments in non-HSVE cases (e.g., brain injury, DAT, stroke and others) reviewed in Capitani et al[2], divided by the overall magnitude of impairment (mild/mod vs mod/severe) and the amount of time elapsed between injury and assessment (short/long gap) The pattern differs qualitatively from that observed in HSVE (see Figure S7).

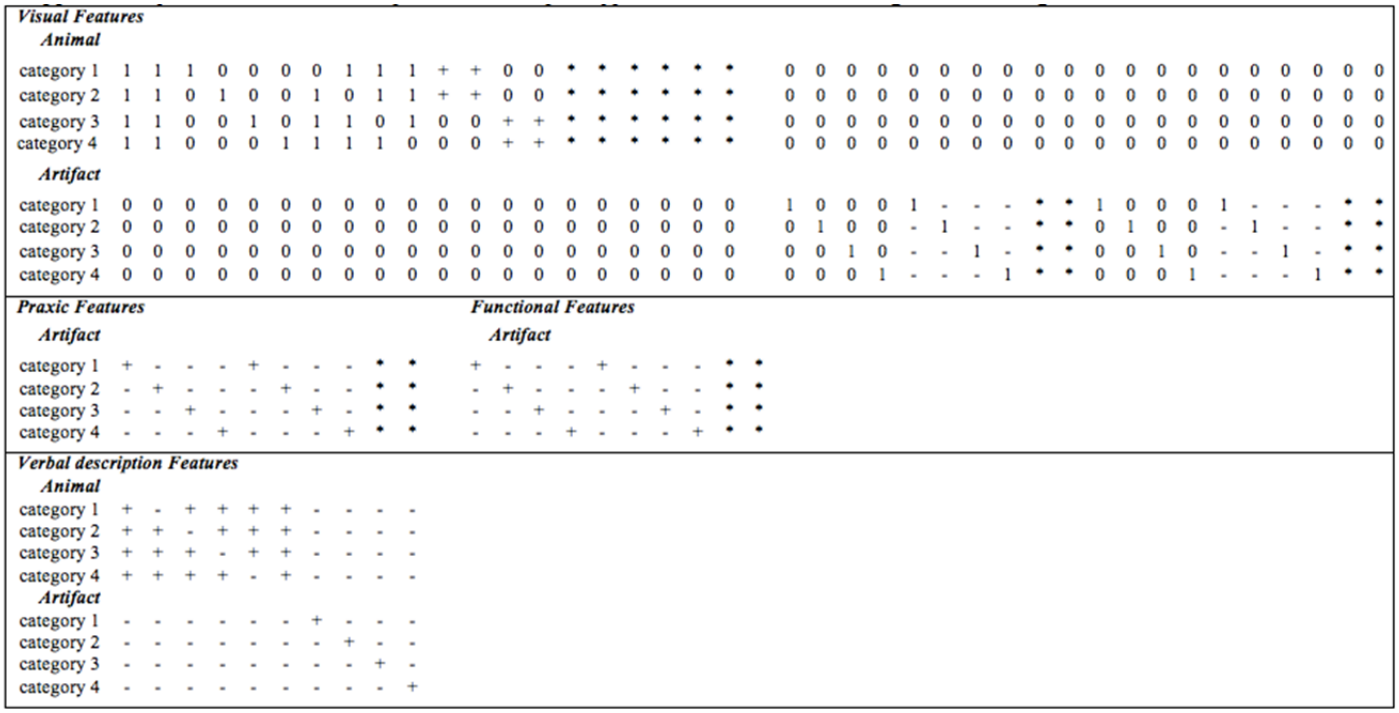
Cluster	Hemisphere	Region (Brodmann's area)	Weighted Center (x, y, z) in MNI space			Volume (mm^3)	ALE Statistics ($\times 10^{-2}$)
Main effect: Animal							
1	Right	Fusiform gyrus (BA 19/37)	45	-62	-15	6256	1.85
2	Left	Fusiform gyrus (BA 37)	-41	-59	-14	2888	1.76
3	Left	Lateral occipital cortex (BA 19)	-42	-81	0	2040	1.89
4	Right	Occipital pole (BA 17)	18	-95	4	952	1.61
5	Left	Occipital pole (BA 17)	-10	-94	11	816	1.15
6	Left	Fusiform gyrus/ITG (BA 36)	-37	-33	-21	632	1.31
Main effect: Artifact							
1	Left	pMTG (BA 37)	-51	-61	1	6064	2.86
2	Left	Fusiform gyrus (BA 37)	-28	-50	-12	4224	3.06
3	Left	Inferior Frontal Gyrus (BA 9)	-49	12	25	3168	2.63
4	Left	LP_superior (BA 40)	-36	-43	48	2928	2.12
5	Right	Fusiform gyrus (BA 40)	-40	-36	46	2800	2.45
6	Right	LP_inferior (BA 13/40)	-58	-28	37	1400	3.35
Conjunction analysis							
1	Left	Fusiform gyrus (BA 37)	-34	-54	-13	616	1.23
2	Left	Fusiform gyrus (BA 37)	-47	-63	-10	272	1.36
3	Right	Fusiform gyrus (BA 37)	40	-54	-16	8	0.86
Contrast analysis: Animal vs. Artifact							Z scores
1	Right	Fusiform gyrus (BA 20/37)	45	-67	-12	3800	3.89
2	Left	Lateral occipital cortex (BA 19)	-40	-82	0	1488	3.54
Contrast analysis: Artifact vs. Animal							Z scores
1	Left	pMTG/ITG (BA 37)	-54	-60	2	2968	3.89
2	Left	LP_superior (BA 40)	-39	-41	47	2184	3.24
3	Right	Fusiform gyrus (BA 37)	27	-50	-12	1984	3.53
4	Left	Fusiform gyrus (BA 37)	-24	-45	-13	1120	2.97
5	Left	LP_inferior (BA 40)	-57	-29	37	920	3.24

Supplementary Table S1. Main effect, conjunction and contrast results from the ALE meta-analysis. The ALE maps for main effects of animal and artifact concepts revealed concordance of peak coordinates across studies that show activation for animal or artifact concepts over baseline conditions. Conjunction and contrast analyses then were conducted on the main effect maps of animal and artifact concepts. Note that animal effect was observed in left pFG in main effect analysis but diminished in contrast analysis. Also, the artifact effect in both main effect and contrast analyses consisted of two clusters in lateral parietal cortex, one more inferior and the other more superior. ITG = inferior temporal gyrus; pMTG = posterior middle temporal gyrus; LP = lateral parietal cortex.

		From (seed regions)					
		ATL	FG_lat	FG_med	MTG	IPL_1	IPL_2
MNI coordinates (x, y, z)		-39, -18, -30	-40, -54, -13	-20, -40, -13	-47, -52, -4	-40, -38, 41	-49, -44, 48
To (target regions)	ATL	**1.000**	**0.095**	**0.131**	0.028	0.000	0.001
	FG_lat	**0.100**	**1.000**	**0.425**	***0.042***	0.001	0.001
	FG_med	**0.149**	**0.428**	**1.000**	0.010	0.000	0.000
	MTG	***0.060***	**0.082**	*0.016*	**1.000**	*0.025*	**0.233**
	IPL_1	0.000	0.000	0.000	0.001	**1.000**	**0.619**
	IPL_2	0.000	0.000	0.000	*0.034*	**0.711**	**1.000**

MNI coordinates (x, y, z)		FG_med	Ant_ILF	Waypoint	SPL_1	SPL_2
		-20, -40, -13	-23,-18,-24	-14, -60, 34	-34, -41, 48	-22, -61, 53
To (target regions)	FG_med	**1.000**	***0.076***	0.008	0.001	0.002
	Ant_ILF	**0.164**	**1.000**	***0.163***	0.001	0.005
	Waypoint	0.011	**0.081**	**1.000**	0.003	**0.126**
	SPL_1	0.000	0.000	0.002	**1.000**	*0.053*
	SPL_2	0.001	0.003	*0.077*	***0.045***	**1.000**

Supplementary Table S2. Connectivity matrix for ventral temporal with both inferior (top) and superior (bottom) parietal dorsal networks. The values in the table are averaged probability (max. streamlines/total no. of streamlines) across 24 subjects from seed ROI regions (in columns) to target ROI regions (in rows). In most of the cases, the matrix is symmetrical qualitatively but not quantitatively. MNI coordinates showed in the table are those used to define spherical ROIs in MNI template. **Bold**: exceeding 5% probability threshold in at least 12 subjects; ***Bold and italic***: exceeding 2.5% probability threshold in at least 12 subjects; *Italic and underlined*: exceeding 1% probability threshold in at least 12 subjects.



Supplementary Table S3. Schematic representations of prototypes of animal and artifact categories in training environment. 1 and 0 denote the absolute probability of 1 and 0 for according features to be present or not; Prob(+) =0.8, Prob(-) = 0.2, and Prob(*) = 0.5. The praxic features of artifacts were created in correspondence to their visual features in order to capture visual affordance between visual and praxic features. Therefore, when a visual feature was present, there was a high probability (p=0.8) for corresponding praxic feature to be present; when the visual feature was not, the corresponding praxic feature had a low probability (p=0.2) to be present. For simplicity, the functional features of artifacts were exact copies of praxic features. For all animal exemplars, all praxic and functional features were turned off (i.e., 0).

	Visual viewing task (sighted)			Name comprehension (sighted)			Name comprehension (blind)		
	Animal	Artifact	*t-value*	Animal	Artifact	*t-value*	Animal	Artifact	*t-value*
Accuracy (%)	97.50 (3.45)	98.33 (2.11)	-0.80	97.50 (3.45)	98.61 (2.03)	-1.07	100.00 (0.00)	99.72 (1.08)	1.00
Activation									
Lateral pFG	0.54 (0.08)	0.27 (0.08)	9.13***	0.55 (0.09)	0.27 (0.07)	9.13***	0.18 (0.05)	0.14 (0.06)	1.76
Medial pFG	0.27 (0.08)	0.48 (0.09)	-5.91***	0.28 (0.09)	0.47 (0.08)	-6.07***	0.07 (0.05)	0.32 (0.04)	-16.22***
pMTG	0.08 (0.02)	0.21 (0.03)	-14.19***	0.08 (0.02)	0.20 (0.03)	-13.67***	0.11 (0.05)	0.26 (0.03)	-9.58***
IPL	0.21 (0.05)	0.41 (0.03)	-13.72***	0.13 (0.04)	0.40 (0.03)	-19.98***	0.16 (0.06)	0.40 (0.03)	-13.92***
SPL	0.12 (0.05)	0.40 (0.03)	-18.94***	0.13 (0.04)	0.40 (0.03)	-21.34***	0.12 (0.03)	0.39 (0.03)	-26.43***

Supplementary Table S4. Performance accuracy and hidden-layer unit activations of both sighted and blind models after 100k training when were tested on all training examples. Paired sample *t* tests showed that (a) all models learned to produce the full patterns of representations with partial knowledge from any modality, and the learning was equally satisfactory across two categories; and (b) demonstrate that the activation patterns in hidden layer captured the observed patterns in corresponding brain regions as revealed by previous studies on both sighted and congenitally blind populations. *** < .001 after Bonferoni correction for multiple comparisons. For Accuracy, adjusted *p* values were α /3; and for activation, adjusted *p* values were α /15.

Severity Level	% Affected connection	SD (accuracy)		HSVE (accuracy)		TPT (accuracy)		% Affected connection	VA (latency)	
		Animal	Artifact	Animal	Artifact	Animal	Artifact		Animal	Artifact
1	10	0.41	0.50	0.98	1.00	0.93	0.90	2.5	0.78	0.16
2	20	0.28	0.35	0.97	0.99	0.89	0.84	5.0	1.36	0.33
3	30	0.22	0.26	0.94	0.99	0.85	0.79	7.5	2.07	0.47
4	40	0.18	0.22	0.89	0.96	0.81	0.76	10.0	2.77	0.59
5	50	0.16	0.20	0.89	0.96	0.79	0.74	12.5	3.17	0.76
6	60	0.14	0.19	0.74	0.91	0.77	0.72	15.0	3.93	0.89
7	70	0.12	0.17	0.60	0.81	0.77	0.71	17.5	4.65	0.98
8	80	0.10	0.16	0.37	0.64	0.77	0.69	20.0	5.18	1.21
9	90	0.09	0.14	0.22	0.29	0.78	0.67	22.5	5.73	1.41
10	100	0.08	0.13	0.00	0.00	0.79	0.62	25.0	6.06	1.41
Mean Accuracy										
Model		0.18	0.23	0.66	0.75	0.82	0.74		3.57	0.82
Patient		0.24	0.28	0.50	0.78	0.74	0.64		4.53	2.86

Supplementary Table S5. Overall performance of sighted models after lesion with different levels of severity as a function of proportion of affected connections.

First Author (year)	Case	Etiology	Exam Gap	Task	Animal	Artifact	Overall_Imp	Difference
Barbarotto et al., (1996)	FA	HSVE	0.08	Naming	0.23	0.60	0.59	0.37
Sirigu et al., (1991)	FB	HSVE	0.17	Naming	0.20	0.50	0.65	0.30
Ferreira et al., (1997)	PR	HSVE	0.17	Naming	0.24	0.53	0.62	0.29
Ferreira et al., (1997)	VG	HSVE	0.21	Naming	0.27	0.78	0.48	0.51
Warrington et al., (1984)	JBR	HSVE	0.33	Naming	0.06	0.67	0.64	0.61
Borgo et al., (2001)	MU	HSVE	0.38	Naming	0.33	0.75	0.46	0.42
Warrington et al., (1984)	SBY	HSVE	0.42	Naming	0.13	0.60	0.64	0.47
Barbarotto et al., (1996)	FI	HSVE	0.42	Naming	0.13	0.57	0.65	0.44
Laiacona et al., (1997)	LF	HSVE	0.50	Naming	0.53	0.87	0.30	0.34
De Renzi et al., (1994)	FELICIA	HSVE	0.58	Naming	0.33	0.90	0.39	0.57
Wilson et al. (1997)	CW97	HSVE	0.67	Naming	0.46	0.83	0.35	0.38
Laiacona et al., (1997)	EA	HSVE	0.67	Naming	0.00	0.17	0.92	0.17
Pietrini et al., (1988)	JV	HSVE	1.00	Naming	0.37	0.77	0.43	0.40
Pietrini et al., (1988)	RM	HSVE	1.50	Naming	0.27	0.47	0.63	0.20
Warrington et al., (1984)	KB	HSVE	2.08	WPM	0.45	0.85	0.35	0.40
Warrington et al., (1984)	ING	HSVE	3.58	WPM	0.80	0.97	0.12	0.17
Gainotti at al., (1996)	LA	HSVE	4.04	Naming	0.10	0.69	0.61	0.59
Moss et al., (1997)	SE	HSVE	7.00	Naming	0.64	0.84	0.26	0.20
Young et al., (1989)	MS	HSVE	10.00	Naming	0.19	0.79	0.51	0.60
Hanley et al. (1989)	BD	HSVE	NA	Naming	0.42	0.83	0.38	0.41
Hanley et al. (1989)	BD	HSVE	NA	Naming	0.67	0.89	0.22	0.22
Sartori et al., (1993a)	DANTE	HSVE	NA					
Dixon et al., (2000)	FS	HSVE	NA	Naming	0.50	1.00	0.25	0.50
Sartori et al., (1993b)	GIULIETTA	HSVE	NA	Naming	0.56	0.88	0.28	0.31
Lecours, et al., (1999)	IL	HSVE	NA	Naming	0.35	0.80	0.43	0.45
Swales, et al., (1992)	JH	HSVE	NA					
Sartori et al., (1993)	Michaelangelo	HSVE	NA	Naming	0.30	0.70	0.50	0.40
Tyler et al., (1997)	RC	HSVE	NA	Naming	0.00	0.46	0.77	0.46
Sheridan et al., (1993)	SB	HSVE	NA	Naming	0.10	0.35	0.78	0.25
Gentileschi et al., (2001)	EMMA	Prog atrophy	0.00	Sem. Ques	0.76	0.93	0.84	0.18
Riddoch et al., (1987a)	HJA	Stroke	0.00	Naming	0.34	0.71	0.53	0.37
Ferreira et al., (1997)	MC	Stroke	0.17	Naming	0.24	0.53	0.39	0.29
Hart et al., (1992)	KR	Paraneoplastic	0.25	Naming	0.50	1.00	0.75	0.50
Hillis et al., (1991)	PS	Injury	0.33	Naming	0.39	0.90	0.64	0.50
Laiacona et al., (1993)	FM	Injury	0.42	Naming	0.20	0.70	0.45	0.50
Carbonnel et al., (1997)	EC	Anoxia	0.42	Sem. Categ	0.45	0.90	0.68	0.45
Laiacona et al., (1993)	GR	Injury	0.58	Naming	0.13	0.73	0.43	0.60
Wilson et al. (1997)	KG	Injury	0.67	Naming	0.29	0.42	0.35	0.13
Samson, et al., (1998)	JENNIFER	Injury	0.92	Naming	0.36	0.78	0.57	0.42
Arguin et al., (1996)	ELM	Stroke	2.71	Naming	0.56	0.88	0.72	0.32
Lambon Ralph et al., (1998)	DB	DAT	3.00	Naming	0.59	0.91	0.75	0.31
Cardebat et al., (1996)	GC	Prog aphasia	3.00	Naming	0.08	0.42	0.25	0.35
De Haan et al., (1992)	NR	Injury	3.92	Naming	0.00	0.50	0.25	0.50
Barbarotto et al., (1995)	MF	Prog atrophy	5.04	Naming	0.37	0.87	0.62	0.50
Caramazza et al., (1998)	EW	Stroke	8.00	Naming	0.41	0.94	0.68	0.53
Wilson et al. (1997)	TS	Injury	14.00	Naming	0.29	0.58	0.44	0.29
Rumiati et al., (1994)	Mr.W	?	NA	NA	NA	NA	NA	NA
Humphreys et al., (1997)	DM97	Abscess	NA	Naming	0.46	0.75	0.61	0.29
Humphreys et al., (1997)	SRB	Bleeding	NA	Naming	0.74	0.97	0.86	0.24
Farah et al., (1992)	TU	Bleeding	NA	NA	NA	NA	NA	NA
Gonnermann et al., (1997)	GP97	DAT	NA	Naming	0.67	0.97	0.82	0.30
Mauri et al., (1994)	HELGA	DAT	NA	Naming	0.44	0.78	0.61	0.34
Funnell (2000)	NA	DAT	NA	NA	NA	NA	NA	NA
Riddoch et al., (1987b)	JB	Injury	NA	NA	NA	NA	NA	NA
Farah et al., (1991)	LH	Injury	NA	Naming	0.52	0.84	0.68	0.32
Farah et al., (1991)	MB	Injury	NA	Naming	0.33	0.77	0.55	0.43
Magnie et al., (1999)	JMC	Post-anoxic	NA	Naming	0.00	0.58	0.29	0.58
Capitani et al. (1993)	CA	Prog aphasia	NA	NA	NA	NA	NA	NA
Basso et al., (1988)	NV	Prog aphasia	NA	Naming	0.00	0.15	0.08	0.15
McCarthy et al., (1988)	TOB	Prog aphasia	NA	Word definition	0.33	0.89	0.61	0.56
Breedin et al., (1994)	DM94	Prog atrophy	NA	Object decision	0.33	0.56	0.44	0.23

Supplementary Table S6. Summary of previous HSVE and non-HSVE single cases reported in Capitani et al[2] indicating the time gap between disease onset and first neuropsychological examination. All single cases reviewed are listed in this table. If the patient was tested with other tasks (e.g., word picture matching) besides the picture naming, only the data from the picture naming task were used. Data from cases whose time gap cannot be inferred were also excluded. Onset = time of diagnosis (years); 1st exam = time of first examination (years); Exam gap = time between onset of disease and first examination; Overall_Imp = overall impairment [1-(Acc_{animal} + $Acc_{artifact}$)/2]; difference = category-specific effect (artifact - animal).

	Estimate	Std. Error	t-value	p-value
Model: HSVE variant				
(Intercept)	1.26E-01	3.23E-02	3.897	<.001
Time Gap	-2.87E-05	1.04E-05	-2.753	0.007
Impairment	-3.38E-02	5.20E-02	-0.650	0.517
Time Gap*Impairment	4.20E-05	1.68E-05	**2.501**	**0.014**
Model: HSVE+ variant				
(Intercept)	2.58E-01	7.84E-02	3.288	0.001
Time Gap	-5.33E-05	2.53E-05	-2.110	0.037
Impairment	-5.11E-02	1.26E-01	-0.404	0.687
Time Gap*Impairment	8.70E-05	4.07E-05	**2.137**	**0.035**
Patient: HSVE cases (n = 19)				
(Intercept)	0.603	0.124	4.849	<.001
Time Gap	-0.119	0.040	-3.006	0.009
Impairment	-0.408	0.222	-1.842	0.085
Time Gap*Impairment	0.272	0.082	**3.298**	**0.005**
Patient: non-HSVE cases (n =17)				
(Intercept)	0.474	0.153	3.105	0.008
Time Gap	-0.034	0.042	-0.801	0.437
Impairment	-0.147	0.267	-0.550	0.591
Time Gap*Impairment	0.067	0.082	0.824	0.425

Supplementary Table S7. Regression analysis for the change of category effect over time in model simulations, HSVE patients and non-HSVE patients.

Supplementary Discussion

1. Motivation for regions of interest identified in prior work

The connectivity analysis employed seed regions for parts of the cortical semantic network identified in prior work from structural and functional brain imaging and computational modeling of healthy and disordered semantic cognition[5–10]. We here briefly review the central findings motivating inclusion of these areas.

1.1 Representations of perceived speech in STG. Perceived speech is thought to be encoded along the rostral-going extent of the superior temporal gyrus/sulcus[11]. Evidence for this view stems from functional brain imaging studies assessing responses to spoken-word stimuli compared to nonword auditory stimuli preserving or eliminating various acoustic speech cues. In such studies, primary auditory cortex responds to all manner of auditory stimuli but responses become more selective to speech and more robust to elimination of lower-level acoustic information as one moves anteriorly along the STG/STS. Such results have been observed across several labs[12–15] and the view that anterior STG encodes representations of spoken words is now widespread[16].

1.2 Visual representations of objects in ventral visual stream. The view that posterior fusiform/IT cortex encodes visual representations of objects dates to the classic work of Goodale and Milner[17] characterizing the ventral visual processing stream, and has since received extensive support from neuroimaging[18,19], neurophysiology in humans[20] and non-human primates [21], and neuropsychological studies of acquired visual disorders[22,23].

1.3 Action representations in left parietal cortex. The important role of left lateral parietal cortex (LP) in supporting action knowledge has its roots in the seminal work of Goodale et al. suggesting that the dorsal visual stream plays a key role in visually-guided action[17,24]. Evidence for this view stems from studies of apraxia, in which parietal pathology can disrupt everyday object-related actions such as posting a letter through a slot[25] or demonstrating the misuse of common tools even while basic motor functioning is preserved[26]. Careful behavioural examination and lesion-symptom mapping in such studies have recently suggested that the dorsal stream may support two different kinds of knowledge about object-directed action[27,28]. More dorsal pathology appears to disrupt praxis, that is, the immediate actions with which tools are engaged, such as how they are grasped or the trajectory with which the hand approaches the object[29,30]. More inferior pathology, in contrast, appears to disrupt knowledge of object function: objects are grasped correctly but are used toward the incorrect ends, or in conjunction with the wrong objects [31]. For instance, given the task of lighting a candle, such patients may pick up the candle and attempt to strike it against the matchbox. Such studies suggest two dorsal streams for visually-guided action, a conclusion bolstered by functional brain imaging studies in healthy participants. As one example, when matching objects on the basis of their praxis (e.g. matching a piano and typewriter because they generate similar actions toward different functions), participants showed elevated activity in dorso-parietal regions, but when matching on the basis of function (e.g. matching the piano and violin because both make music despite generating different praxis), elevated activity was observed in more inferior parietal regions[32].

Further evidence comes from neuroimaging studies investigating the neural basis of tool cognition. Tools elicit more activation than other objects in several regions, but the most consistently reported has been the left inferior parietal lobule (IPL)[32,33]. Participants asked to assess a tool's functions or related actions show elevated activation in the left temporoparietal junction, especially IPL[32,34–36]. Praxis-related activation has also been observed in the more superior aspect of left lateral parietal cortex, both in individual studies[37–40] and prior meta-

analyses[41,42]. From these results, our connectivity analysis includes seed regions in inferior and superior aspects of lateral parietal cortex, respectively.

1.4 Anterior temporal lobe as a cross-modal semantic "hub." The importance of the anterior temporal lobe (ATL) for semantic representation has been established through a variety of methods[43,44]. Neuropsychological studies have shown that atrophy of the anterior temporal regions in some forms of dementia produces a profound disruption of semantic memory that affects all semantic domains across all modalities of reception and expression[45–47]. The impairments are not attributable to widespread cortical pathology, since (a) they are not observed in the early stage of more common and widespread forms of dementia such as Alzheimer's disease[48] and (b) lesion-symptom correlations indicate that the semantic impairments are best predicted by hypometabolism in the ventral aspects of the ATL[49]. Semantic impairments are also observed from other forms of pathology to the ATL, including from herpes viral encephalitis (HVSE)[50] and, more subtly, from ATL resection to remediate epilepsy[51].

For many years, these patient data conflicted with the results of functional brain imaging studies, which rarely showed significant activation of ATL regions in semantic tasks. This discrepancy arose from a range of unfortunate methodological factors[52]: (a) Magnetic field inhomogeneities in ventral ATL (caused by their proximity to air-filled sinuses) substantially degrade and distort the fMRI signal; (b) several PET and fMRI studies failed to include the area in their field of view[52]; and (c) ATL areas are more likely to be identified if semantic performance is contrasted with an active, non-semantic baseline activity. Low-level 'resting' baseline conditions engage the ATL[42] and other regions probably because during 'rest' participants engage in semantically-dependent tasks including remembering, thinking, planning, daydreaming, etc.[53]. Studies that address these issues[54–56], reliably observe semantically-related ventral ATL activation that appears to be equally strong for all conceptual domains and modalities. Such responses have recently been observed very directly through human electrocorticography finding ventral ATL responses to both spoken and written words and to pictures, and further showing that the evoked responses carry information about the semantic classes to which the eliciting stimuli belong[57,58]. Finally, the causal role of ATL regions in semantic processing has been established using transcranial magnetic stimulation in healthy participants and direct cortical stimulation in the grid electrode studies. Such stimulation causes slowing of responses in semantic tasks such as synonym judgment but not in equally challenging non-semantic tasks such as number-judgment[59,60]. Stimulation slows responses equally for both abstract and concrete words[61], and for words denoting animals and artifacts[59], again consistent with the view that the ATL plays a domain-general role in semantic representation.

1.5 Connectivity patterns of ATL, STG, pFG and LP. Anatomical connectivity between the ventral ATL "hub" and regions that encode auditory speech (STG) and visual object (pFG) representations has been documented in both human and animal studies[62,63]. Recent probabilistic tractography with human subjects showed robust connectivity of ventral ATL with pFG and with anterior STG[64]. Thus the anatomical connectivity of the semantic hub (ATL) with visual object and spoken word representations is well documented. To our knowledge, the further question whether medial and lateral aspects of pFG show similar or different patterns of connectivity with ATL (and other intra-temporal semantic regions) has not been investigated prior to the current work.

The connectivity of lateral parietal action representations with the temporal-lobe regions of interest is also not well understood. Primate studies have found long projections from lateral parietal to posterior ventral temporal cortex[65,66], but to our knowledge no direct connection from

lateral parietal to anterior temporal regions has been reported, and in any case it is not clear how well the human and non-human white-matter anatomy aligns. Studies from non-human tracking studies[66] and also from in-vivo white-matter stimulation in humans[63,64] have begun to tease apart different tracts within the fasciculi that traverse the length of the temporal lobes[67]. Most notably, the inferior longitudinal fasciculus, which begins in the ventral ATL, runs inferiorly to ventral occipito-temporal territory, and on to occipital cortex proper. This tract appears to branch dorsally near the pFG to terminate in more dorsal parietal cortex. A pathway from medial posterior ventral temporal to parietal cortex was also reported using deterministic diffusion-weighted tractography[63], though these results are difficult to interpret given the limitations of that method. Similarly, Mahon has reported significant functional connectivity between medial pFG and lateral parietal cortex, including a more inferior region and a more superior region[33,68,69], but the anatomical pathways mediating such relationships remain unclear. Thus the evidence prior to the current work, albeit limited in humans, suggests potential connectivity from posterior ventro-temporal to parietal-lobe action representations that may be more pronounced in medial pFG. A central goal of the tractography reported in the main paper was to measure white matter connectivity amongst temporal and parietal regions of interest in humans using state of the art probabilistic diffusion-imaging methods and techniques for resolving signal-distortion problems in the ventral ATL[52,56].

2. Relationship of current proposal to prior neuro-computational models

Research in human semantic representation and its disorders has been a focus of neuro-cognitive modeling work for many years[9,70]. The current work synthesizes many insights from these prior efforts, but also differs from past work in important respects. To make clear these relationships we briefly review milestones from previous research and note similarities and differences to proposal in the main text.

2.1 Category-specific semantic impairment (*Farah and McClelland, 1991).* To understand how category-specific double dissociations might arise within a system that employs distributed semantic representations, Farah and McClelland[71] investigated a recurrent network implementation of the sensory/functional hypothesis[72,73] articulated earlier by Warrington and Shallice[72,74]. The model proposed feature-based semantic representations in which each unit stood for a particular property (such as has stripes or used to cut). Individual concepts were cast as patterns of activation over these units (i.e., units denoting the item's features received a high activation while those not true of the item received a low activation). Semantic features were grouped based on the kind of information they encoded, and units within a group were assumed to be anatomically close together in the brain and thus more likely to be damaged together in brain injury. The groups included sensory semantic features (referring mainly to the visually apparent properties of objects) or functional semantic features (referring to their use). These semantic representations interacted recurrently with distributed visual representations of objects coded in a separate network layer, and with distributed representations of words in a third layer. The semantic layer thus served as an intermediating structure between visual and phonological representations similar to the current proposal; however semantic knowledge about an item's properties was encoded within the semantic representation itself, and not through recurrent interactions with other modality-specific representations distributed throughout cortex.

On the basis of dictionary definitions of objects, the authors showed that manmade objects typically have more functional and fewer perceptual properties than animals. In computer simulations, they then showed how category-specific deficits could arise within a recurrent

network that employed semantic representations so structured. Specifically, damage to sensory semantic properties disproportionately affected animal concepts, since animal representations relied heavily on these properties and their loss disrupted correct activation of even functional and features (and names) via recurrent interactions. When damage was limited to functional features, knowledge of manmade objects was disproportionately affected. The model provided the first demonstration of how category-specific double-dissociations could arise in a system that did not dedicate separate representational modules for the dissociated categories.

The Farah and McClelland model clearly resonates with the current proposal in suggesting that category-specific dissociations arise partly from the differential reliance of animal and artifact concepts on visual versus praxis/functional information. Yet it differs in several respects: (1) Semantic representations were assigned by the theorists, and not learned through cross-modal mappings, raising the question of where the semantic information comes from and how it is acquired. (2) In using explicit feature-based semantic representations, the model separates perceptual, linguistic, and semantic information in ways that make it difficult to understand how semantic meanings are grounded[75–77]. (3) The model was intended to explain category-specific double dissociations, and it is not clear how domain-general impairments of the kind observed in SD might arise within a system where semantic features are partitioned by the kind of information they encode. Subsequent work with the same model showed that perfuse damage across all semantic features could produce category-specific effects that would change direction depending upon the magnitude of the damage[78], but this pattern is not observed in SD[48,79]. (4) The model was not anatomically constrained, and the authors made no claims about how its architectural components might relate to real brain anatomy, apart from the suggestion that semantic features of a given kind might be co-localized in the brain. (5) It was not clear what constitutes a functional feature or how these might be localized; for instance, the ability to fly is sometimes cited as a "functional" feature of birds, but this is clearly very different from knowledge about the praxis with which objects are engaged or the uses to which they are put. (6) It is not clear how such a model might explain category-specificity in functional brain imaging studies of the congentially blind, since category effects largely depend on knowledge about visual properties of animals.

2.2 Connectivity and functional specialization (*Lambon Ralph et al., 2001* and *Plaut, 2002*). Two early models investigated how differential connectivity in neural networks might give rise to graded functional specialization, as reflected in patterns of impairment following brain damage. First, Lambon Ralph et al.[80] used a model similar to Farah and McClelland[71] to investigate lateralization of function in semantically-impaired patients. The model assumed that the phonological/word representations in this model were encoded in the left hemisphere; that feature-based semantic representations and visual representations were bilaterally distributed; and that within-hemisphere connections were more robust than cross-hemisphere connections. These assumptions were implemented by dividing the semantic and visual representations into two groups, one for each hemisphere, and setting the model learning rate higher (more effective) for connections within a hemisphere than for connections between hemispheres. Because word representations were assumed to be left-lateralized, the ability to produce correct name output from a visual image or a specific semantic representation depended more upon semantic features encoded in the "left" part of the model than those in the "right" part of the model. Yet because semantic features were distributed bilaterally, the ability to comprehend a word or image was equally impaired by left-lateralized or right-lateralized pathology in the semantic layer. The model thus predicted that left-predominant pathology should affect verbal production more than

comprehension, but that production and comprehension should be more equally affected for right-predominant pathology—a pattern subsequently documented in semantic dementia[80]. In relation to the current work, this model inherits the same points of contrast noted above for the Farah and McClelland model. The model was thus the first to show how connectivity constraints could produce a graded impact on the magnitude of deficits observed across tasks (in this case, verbal production versus nonverbal comprehension), but did not otherwise address questions about category-specificity, neural connectivity in real brains, or different semantic syndromes that are the focus of the current work.

The second model[81] illustrated how connectivity constraints might explain optic aphasia, a puzzling neuropsychological syndrome in which (1) language comprehension is intact, (2) visual recognition appears intact when the patient is assessed with gesture (i.e. demonstrating object use) but (3) is degraded when assessed with language (i.e. when naming an object). To explain this pattern, Plaut[81] proposed that the neural systems that intermediate between visual, praxic, and word representations are subject to an evolutionary constraint that favors short connections in the brain. Under this view, neurons that are anatomically proximal are more likely to mutually influence one another. Thus neurons that are closer to visual and action systems end up, through learning, contributing more to the ability to map from visual input to action; neurons lying closer to visual and phonological systems contribute more to the mapping from vision to speaking; and so one. Neurons that are anatomically equidistant to these modality-specific representations contribute equally to all mappings. Plaut showed that the symptoms of optic aphasia arise in such a system when damage targets unit lying between vision and naming but spares those lying between vision and action.

The optic aphasia model was the first to show how graded modality-specificity could arise through learning and connectivity operating together in the semantic system. The current work extends this idea in several important ways. First, we show that graded connectivity can produce category specificity as well as domain-general impairments—patterns critical for theories of semantic memory not addressed in this early work. Second, whereas the early work proposed a general constraint preferring local connectivity, the current work shows that long-range white-matter connectivity plays a critical role in the organization of the semantic network. Third, the author did not consider how the model architecture relates to real brain anatomy. Fourth, the model focused solely on optic aphasia—it was not brought to bear on category-specific double dissociations, domain-general semantic impairments, or patterns of functional activation in brain imaging studies.

2.3 Computational arguments for a semantic hub (*Hinton, 1986; Rumelhart, 1990; Rogers and McClelland, 2003, 2004, 2005, 2008*). Our proposal has roots in important computational work by Hinton[82] and Rumelhart[83,84], who were the first to show how distributed semantic representations could emerge through learning in the internal (hidden) layers of a neural network model trained to report the features of familiar items in the environment. Hinton's classic "family-trees" paper showed that such representations could express quite abstract similarity structure that provided a basis for knowledge generalization. Rumelhart showed that the same principles explain how propositional semantic knowledge could be encoded in the weights of a feed-forward network that learned distributed internal representations. In a book and a series of papers, Rogers and McClelland[7,75,76,85] argued that Rumelhart's model offered a new, unified account of several previously puzzling phenomena in the study of semantic cognition, including results from studies of semantic knowledge in infancy and childhood, healthy adulthood, and in semantic dementia.

A key contribution of this work was the demonstration that many of the model's interesting properties would only emerge in a convergent architecture—that is, within a network where there exists at least one layer that contributes to representation and processing of all kinds of items, across all semantic tasks. When different parts of the network are "dedicated" to only a subset of items or tasks, the system becomes insensitive to patterns of covariance across items and tasks, and loses the ability to discern "deep" conceptual structure that is only encoded in such covariance. This observation provided the central computational motivation for the proposal that the semantic system requires a cross-modal "hub" that contributes to representations for all kinds of concepts, across all receptive and expressive modalities. This work also provided the first demonstration that a central characteristic of the impairment in semantic dementia—specifically, the preservation of knowledge about more general or superordinate category properties relative to more specific or subordinate categories—arises from damage to networks that acquire distributed internal semantic representations through learning.

The current model conforms to the "convergence" principle articulated in this work in adopting a cross-modal semantic hub important for all kinds of concepts. Also, as in this early work, semantic dementia is proposed to arise from damage to a central semantic representation. Otherwise the work makes very different contributions. Apart from proposing a cross-modal hub, the early model did not stake claims about brain anatomy, did not consider how category-specific patterns might arise from brain damage, did not consider functional or structural brain imaging results, and did not advance specific proposals about how the content the network encoded might relate to sensory, motor, linguistic, and affective systems in the brain.

2.4 The hub-and-spoke model (*Rogers et al. 2004; Lambon Ralph et al. 2007; Patterson, Nestor and Rogers, 2007).* Building on detailed study of the behavioral impairments and neuropathology observed in SD[79], together with the computational insights derived from the Rumelhart model, Rogers et al[75] proposed that the anterior temporal lobes function as a "semantic hub" that acquires, through learning, distributed representations that serve two important functions. First, they promote cross-modal interactions amongst various modality-specific sensory, motor, and linguistic representations, permitting inferences about an item's unobserved properties from its visual appearance, its name, or verbal statements about it. Second, the hub was hypothesized to encode conceptual similarity structure amongst items, thus promoting generalization of learning across conceptually similar items, even if these happened to differ in their visual appearance or in other superficial respects.

Despite sharing a similar architecture, the "semantic hub" model departed from the Farah/McClelland model in an important respect: it dispensed with feature-based semantic representations. The "hub" representations were proposed to encode conceptual structure in an inchoate form: conceptually similar items evoked similar patterns of activation in the hub, but these patterns did not encode explicit knowledge of the item's properties. Instead, properties were held to be encoded within modality-specific systems for perception, action, and language. For instance, knowledge that a stop sign is red inheres (on this view) partly in the ability to generate activation within or near parts of cortex that directly encode perception of the color red, and also partly in the ability to generate, in language production systems, the word "red" when the system is asked to verbally report the color. The ability to generate appropriate responses within modality-specific systems was held to be dependent upon the cross-modal semantic hub. For instance, a visual input depicting an object's shape would generate a pattern of activation within the hub, which would then "broadcast" back to other modality-specific systems to generate activation patterns corresponding to the item's name, its characteristic pattern of

motion, associated actions, and so on. Thus the model was able to side-step difficult questions about which of an item's properties "count" as semantic features and which correspond to "mere" perceptual features or verbal labels. It also connected neural-network models of semantics to the emerging view that conceptual meanings are grounded in modality-specific sensory and motor systems[86,87].

The implemented hub-and-spokes model investigated the potential of these ideas to explain patterns of semantic impairment observed in SD[5,45,46,75]. Like the Farah & McClelland model, this was a recurrent neural network in which a "semantic" layer, held to be located in the ATL, mediated interactions between visual representations of objects (held to be encoded in infero-temporal cortex) and representations of verbal statements about objects, including their names (held to be encoded within STG). The model was trained on patterns that expressed similarities amongst various animals and manmade objects, as assessed by verbal attribute-listing studies and a new study of visual feature overlap. Semantic dementia was then simulated by removing an increasing proportion of weights from the model "semantic" layer. The resulting simulations accounted for several detailed aspects of the deficits observed in semantic dementia, and also made several predictions about the role of the anterior temporal lobes in semantic processing, subsequently borne out in further work (for reviews, see Lambon Ralph et al.[44,88,89]). In a subsequent paper, Lambon Ralph, Lowe and Rogers[5] showed how different kinds of damage in the model hidden layer could produce qualitatively different patterns of deficit that resembled the differences observed between SD and patients with semantic impairment from HSVE.

In proposing a cross modal hub, situating the hub and the visual and verbal spokes in specific brain areas, and using the resulting model to account for two neuropsychological syndromes, the hub-and-spoke model was an important precursor to the current proposal. The new work builds on these ideas but also suggests some important additional factors not previously considered. Specifically: (1) The earlier model did not consider how representations of action and function might interact with or otherwise influence behavior of the model. Though schematic figures depicted such representations as interacting directly with the ATL hub[5,59], such interactions were not implemented in the model. (2) The model was assumed to be fully connected—differences in connectivity between different spokes and the hub, and the possibility of interconnection amongst some spokes, were not considered. (3) The model did not consider other cortical regions that might be included in the semantic network. (4) The account of category-specific impairment in HSVE in that model was somewhat different than the proposal from the current model, and the reverse category-specific pattern (worse knowledge of artifacts than animals) was not considered at all, promoting the standoff between domain-general and domain-specific theories of semantic representation that motivates the current work. (5) It was not clear how the model might account for functional imaging data showing category-specific patterns in either healthy individuals or in the congenitally blind, since all units were equally involved in representing all concepts in this model. (6) While the model was consistent with some known aspects of neural connectivity, no effort was made to explore neural connectivity amongst model regions of interest or to relate such investigations to the model architecture.

2.5 Incorporating praxic representations into the semantic network (*Chen & Rogers 2015*). Most recently Chen and Rogers[6] investigated how praxic representations might be incorporated into the hub-and-spokes framework. The paper explored three hypothetical model architectures that differed only in their assumptions about how representations of object-directed praxis/action might interact with visual, verbal, and semantic representations in the hub-and-spoke framework. The first model proposed that such representations connected directly to the

semantic hub; the second, building on functional (but not anatomical) connectivity studies by Mahon and colleagues[33,90], proposed an additional "direct" connection between action representations and visual object representations; the third proposed only the direct vision-to-action connection, without direct connection between action representations and the hub. Consistent with Mahon's suggestion (and the current work), the vision-to-action connections in models 2 and 3 were graded so that more medial parts of the infero-temporal visual system were more strongly connected to action representations than were the lateral aspects. The authors considered the capacity of each model to explain two unrelated phenomena: (1) category-sensitive patterns of functional activation in the pFG of sighted and blind individuals, and (2) the form of apraxia observed in semantic dementia, where the ability to use familiar objects is seriously impaired but the ability to use novel tools to solve mechanical puzzles is normal. The full range of results was only explained by one model architecture—specifically, model 3, where visual and action representations were connected via an anatomically graded direct path, and action representations were disconnected from the semantic hub.

This work raised several critical questions that are answered in the current work. First, it was not clear whether the hypothetical connectivity employed in the successful model is actually observed in real brain anatomy—the tractography in the current paper lays out white-matter tracts connecting the full cortical semantic network. Second, the paper focused solely upon explaining functional brain imaging patterns in the pFG—it was not clear whether category-specific patterns throughout the semantic network could be understood as arising from underlying connectivity, or whether the observed effects would persist in a more complete network expressing real neural connectivity. Indeed, it was not clear where such effects are reliably observed, or what the anatomical connectivity amongst implicated regions actually is. The new ALE analysis in the current work both establishes where category effects are reliably observed, and how the corresponding regions connect in anatomy. Third, the literature[28,91] suggests that functional and praxic representations may be encoded in different lateral-parietal regions, as previously noted. The architecture of the current model is motivated by measured anatomical connectivity between temporal lobe regions and both IPL and SPL. Fourth, the patient simulation work focused solely upon apraxia in semantic dementia—the paper did not consider category-specific impairments in other patient groups, the absence of such effects in SD, or the anatomical bases of the different syndromes. The core contribution of the current proposal—a reconciliation of domain-general and domain-specific views of semantic representation—arose from an effort to provide answers to each of these questions within a single project that marries functional imaging, structural imaging, modeling, and neuropsychology.

3. Simulating different disorders of semantic representation.

We considered model simulations of four disorders of semantic representation, constituting all the acquired disorders that (a) are typically attributed to degraded knowledge/representation, rather than degraded retrieval/access to semantic knowledge, and (b) have been shown in case-series studies to produce consistent patterns of impaired performance on semantic tasks. We here note the simulation of model pathology for each disorder and its motivation.

3.1 Semantic dementia (SD). SD is a neurodegenerative disorder associated with gradual thinning of cortical grey matter and associated white-matter fibers, centered in the anterior temporal lobe[49]. This pathology seriously degrades semantic knowledge for all kinds of concepts,

across all modalities of reception and expression[45,75,92], while generally sparing many other cognitive functions. Because SD is progressive, any relearning or reorganization within diseased regions is continually compromised by later disease. Patients with SD present with word-finding difficulties and verbal comprehension deficits, but detailed assessment invariably reveals pan-modal knowledge impairments, including loss of knowledge about the visual structure and colors of everyday objects, their functions, associated sounds, typical patterns of movement, and even characteristic odors[43]. The loss of semantic knowledge in patients with SD first appears in the idiosyncratic properties that differentiate closely related items, but gradually manifests in properties that distinguish basic-level categories and eventually even those characterizing very general categories. To capture the loss of neurons and white-matter in ATL without benefit of relearning we simulated SD by removing an increasing proportion of all weights entering, leaving, or internal to the ATL hidden layer. Weights were removed in increments of 10% for each of 15 trained models differing only in the random configuration of their initial weights. At each level of damage, the model was tested without allowing it to relearn/reorganize. Reported results are averaged across the 15 different models.

3.2 Herpes Simplex Viral Encephalitis (HSVE). HSVE is a disease that produces rapid bilateral necrosis of gray and white matter. In patients with semantic impairments from HSVE, the pathology encompasses largely the same regions affected in SD, but typically with a greater density of damage in the medial temporal lobes/hippocampus as well as damage to frontal cortex, and temporal white matter especially in the lateral axis[50]. Despite this generally greater extent of pathology, such patients often show less semantic impairment overall, with greater deficits of knowledge for animals than for manmade objects[5,50]. The contrast of SD and HSVE thus creates a puzzle: if pathology in HSVE includes the regions damaged in SD plus additional regions, why is the semantic impairment milder in HSVE, and why is knowledge of artifacts relatively spared?

The main paper considers two potential explanations that are not mutually exclusive. Each is associated with a different model of HSVE pathology. The model first captures differences in the time-course of SD vs. HSVE: whereas the former progresses slowly over a course of years[79], the latter develops rapidly and is then halted by anti-viral medication, followed by some degree of recovery in most patients. Patients with HSVE are typically assessed months or years after the comparatively acute insult, so that the damaged system has had some amount of time to relearn/reorganize. To capture this difference between disorders, the first variant simulated HSVE pathology with the same lesion procedure used for SD (removal of a proportion of weights entering, leaving, or within ATL), but the damaged model was then retrained in the same learning environment and with the same learning rate. Semantic task performance was assessed every 500 epochs through a total of 5000 epochs of relearning, yielding the relearning curves shown in Figure S4. A time point midway through relearning is shown in the main text Figure 4B. The second HSVE variant considered how asymmetric white-matter damage across the lateral/medial axis of the ATL might affect category-specificity in HSVE. Specifically, Noppeney et al.[50] reported a greater extent of white matter damage in HSVE than in SD that was especially pronounced laterally in temporal lobe. To capture this difference, we again employed the same lesion procedure used for SD and HSVE models in ATL units, but with connections between ATL and lateral FG having a greater probability of removal (see Methods). The damaged models were again retrained in the usual learning environment, with semantic task performance assessed every 500 epochs for a total of 5000 epochs as shown in Figure S5.

3.3 Temporo-parietal tumor resection (TPT). The category-specific pattern in HSVE is much more commonly observed than the reverse dissociation of worse knowledge for manmade

objects than animals. The anatomical bases of artifact-specific impairment was unclear for many years, since the relatively small number of patients exhibiting this pattern typically had quite wide-spread lesions[2]. The first relatively large-scale case-series study of this pattern was conducted by Campanella and colleagues[93], who analyzed lesion-symptom correlations in a group of 30 patients who had undergone surgical removal of temporal-lobe tumors (20 in the left and 10 in the right). The group exhibited significantly worse knowledge of nonliving things compared to animals, with difference scores in naming accuracy ranging from 2%-21%. The category effect was most evident for resection of the left posterior temporal lobe, and voxel-based lesion-symptom mapping (VLSM) 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 an increasing proportion of connections between and within IP and MTG model regions.

3.4 Category-specific visual agnosia (VA). Finally, a long tradition of research suggests that forms of associative visual agnosia arising from damage to occipitotemporal regions can have a greater impact on recognition of living than nonliving things[94,95]. The deficit is specific to vision: such patients can access semantic knowledge from other modalities including language. This pattern was recently documented in a case-series analysis of patients diagnosed as "letter-by-letter" readers[96] —an acquired form of dyslexia thought to reflect low-level deficits of visual perception[95]. Consistent with this hypothesis, the group was shown to have difficulty discriminating visual gratings, especially in higher spatial frequency ranges. When assessed on a standard picture-naming task using line-drawings of common animals and objects matched for familiarity, the group showed impairments relative to healthy controls, with significantly worse performance for animals than for manmade objects[96]. For milder patients, the impairment was reflected in naming response times but not accuracy; for more severe patients, the category effect was observed in accuracy as well. The study suggests that subtle impairments of visual perception can produce a category-specific visual recognition impairment, as suggested previously by many groups[97–99]. 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).

4. Assessing how category effects change over time in HSVE and other disorders.

Our account of category-specific impairment in HSVE suggests that the magnitude of the category-specific pattern can change over time with relearning, an important question that has not been explicitly tested in extant literature. Two previous studies[3,4] have examined the category-specific impairment in HSVE patients longitudinally. All patients showed substantial recovery of semantic knowledge on the 2nd examination compared to the 1st examination, but change in the size of the category-specific effect was inconsistent across patients: Two cases showed a reduced category effect at the second session, while the remaining two showed a larger category effect. We therefore conducted simulations and further analysis of the case-study literature to assess (a) whether the model explains these different patterns in the longitudinal studies, (b) whether cross-sectional data at the group level shows a similar pattern to these few longitudinal studies, (c) which factors contribute to the direction of change of category effect, and (d) whether this pattern of change over time is unique to HSVE.

As shown in Figures S4 and S5, the direction of change in the magnitude of the category effect—whether it increases or decreases—depends upon the magnitude of the initial semantic impairment in the model. For milder initial deficits, a category-specific impairment favoring

artifacts diminished as knowledge recovered in both categories. For more severe initial impairments, the category effect increased as knowledge recovered more rapidly for artifacts than for animals. The interaction is not attributable to simple floor/ceiling effects—the interaction is observed even when the model is not completely at floor or ceiling. The model thus predicts that category effects should grow in more severely impaired patients over time, but should shrink in milder patients over time—a pattern consistent with the two longitudinal studies just mentioned (see Figure S7 D).

To further test the model prediction we considered data from a large group of patients with HSVE resulting in category-specific impairment reviewed by Capitani et al.[2]. For each case we computed (a) the time passed between injury and assessment, (b) the overall magnitude of the impairment and (c) the size of the category effect. These data are shown in Table S6. As reported in the main text, a regression on these data showed that the direction and magnitude of change in the category effect was reliably predicted by the interaction of the severity of the initial impairment and the time elapsed between injury and test, in just the manner predicted by the model: category effects reduced over time in milder patients but grew over time in more severe patients ($t = 3.298$, $p < .01$) Finally, to assess whether this pattern was common to all varieties of semantic impairment, we conducted the same analysis on patients with semantic impairment from other etiologies (including brain injury, stroke, DAT and other progressive pathologies) also reviewed by Capitani et al.[2]. In contrast to the HSVE cohort, change in the magnitude of the category effect was not reliably predicted by the magnitude of the initial deficit, the time elapse between injury and test, or their interaction (Figure SI-8 and Table SI-7).

Supplementary Methods

1. Study selection process for ALE analysis

In the ALE analysis we investigated where category-sensitive effects are reliably observed in functional brain imaging studies that employ either words or pictures as stimuli, and whether such studies implicate additional cortical regions beyond those included identified in previous neuro-computational models of semantic representation. We first searched PubMed for articles using the terms "category-specific", "living", "nonliving", "tool" and "animal" in combination with "fMRI", "PET" or "neuroimaging" in either the title or abstract up to July 2013. To this search, we added studies cited in three literature reviews [1,100,101] and three meta-analyses [41,102,103]. The first author (L.C.) examined these papers and filtered them using five criteria. Specifically, the studies (a) had to use data from healthy human participants, (b) had to employ pictures or words denoting animals and/or artifacts (excluding faces, body parts, and landmarks), (c) had to report activations as Talairach or MNI coordinates from univariate analyses, (d) excluded reviews, large-scale study of data re-analysis, and studies using other techniques such as EEG or TMS, and (e) excluded studies of social categories. With this approach we found 49 papers describing 73 independent studies (31 for animal and 42 for artifact) and reporting a total of 270 foci (103 for animal and 167 for artifact). Papers and the resulting table of foci are listed in online materials downloadable from the following sites (https://github.com/halleycl/ChenETAL_NatHumanBehav_SI-Online-materials and https://app.box.com/v/ChenETAL-NatHumanBehav-SI).

2. ROI definition of probabilistic tractography

All seed regions were in the left hemisphere and we restricted our analysis to this hemisphere. Coordinates of peak activation from the meta-analysis and the imaging study in MNI brain space[104] were used to define the ROIs. Seeds were placed in the white matter underlying cortical peaks based on the group-averaged ACM map. Because the medial pFG group coordinate was on the edge of the ventral occipito-temporal cortex and very close to cerebellum, the seed was placed in temporal white matter using each individual's ACM map. A sphere with a diameter of 6mm centered on the seed coordinate for each ROI was then drawn in the MNI template. Finally, the ROIs defined in a common space were converted into the native brain space of each individual.

3. Thresholding and group-average of tractography maps

On 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. At each individual level, a 2.5% threshold was applied to all cell values so that the connection is considered to be highly probable only when the connectivity value exceeded 2.5% of the total number of generated streamlines (i.e., 2.5% of 15,000). Since the criteria 2.5% was chosen heuristically based on a previous study[105], we also considered a more stringent criterion of 5% and a more lenient criterion of 1% 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. To allow for anatomical localization and inter-subject comparisons, the tracking results after the 2.5% threshold for each participant were spatially normalized to the

MNI template space using the DARTEL toolbox supplied as part of SPM8[106]. A group-averaged tractography image was then obtained by averaging the normalized individual data.

4. Spatial gradient of visual-praxic connections in model simulations

To capture the observation that medial pFG is more strongly connected to parietal regions than is lateral pFG[33,68], the 20 units in the pFG layer were situated along an anatomical lateral-to-medial axis. The learning rate on visuo-praxic connections was scaled according to the visual unit's position along this axis, with medial-most units having a larger learning rate than other connections in the model, the rate diminishing for increasingly lateral units, and the lateral-most units having a smaller learning rate than other connections. More specifically, the error derivatives (and hence the strength of influence on weight changes) on visuo-praxic connections was scaled according to the visual unit's position along this axis, with medial-most units having a larger scaling factor than other connections in the model and the rate diminishing for increasingly lateral units according to the following sigmoid function:

$$\boldsymbol{S(i) = 2/(1 + e^{\frac{n}{2} - i})}$$

where S(i) is the scaling parameter applied on error derivatives, n is the number of units in vision layer, and i (the unit location) ranges from 1 to 20 on the medial-lateral axis. Across units the mean learning rate was equal to that on other connections in the model. As a result, the lateral-most unit has an error derivative value close to zero, whereas the medial-most unit has an error derivative almost double the magnitude of other model connections.

5. Training representations and procedures for model simulations

A model environment was constructed to contain visual, verbal, function/action and praxic representations for 24 different exemplars of animals and 24 different exemplars of tools, with each domain organized into 4 basic categories, each containing 6 exemplars (for schematic prototypes, see Supplementary Table S3). 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.[75] in their analysis of verbal attribute-listing norms and line drawings of objects. Thus (a) items in different domains shared few properties; (b) items within the same category shared many properties; (c) animals from different categories shared more properties than did artifacts from different categories; and (d) animals had more properties overall than did artifacts. Each item was also given a unique name as a 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 praxic layer. For all animal items, these units were turned off, capturing the general intuition that most animals are not associated with rich praxis. For artifacts, the states of ten visual units were directly copied for each item, and each feature in each item representation was then flipped with small probability ($p = 0.2$) to create distorted versions of the visual pattern. This approach ensured that praxis patterns, while not identical to the visual patterns, still captured the relevant category structure. It also provided a model analog to visual affordance: particular visual and praxic features occurred together with high but non-certain probability across items. For simplicity, the function representations duplicated the praxic patterns across the 10 visible units for function features.

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 praxic features. Model inputs could be from just one modality of the following: single names (one

verbal unit activated), verbal descriptions (multiple verbal units activated), visual images (visual features activated), functions (functional features activated) or praxis (praxic features activated). Inputs were applied to visible units by providing these with direct excitatory input, and units throughout the network were updated successively over time in random order. After 8 update cycles, target values were applied to all visible units, indicating the item's all properties including visual, verbal name, and verbal description properties, and for artifacts, function and praxic properties as well. 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[107]. 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.

The response to a given input was counted as correct if the pattern generated across all visible units was on the correct side of the activation midpoint (i.e., target properties were active above 0.5, non-target properties had activation below 0.5). All models were trained exhaustively for 100k epochs at which point they generated correct output by this criterion across all visible units for the great majority (>94%) of inputs.

6. Assessing the change of category effect over time in HSVE and other disorders

For model simulations, naming accuracy from both HSVE and HSVE+ variants with different relearning was recorded. The amount of relearning (from 0.5k epochs to 5k epochs) was taken as the analog of time between injury and first assessment in the patient data. The proportion of connections removed was taken as a model analog of the initial magnitude of impairment. The category effect was measured as the difference in naming accuracy for artifacts versus animals. Regression analysis was conducted to test whether time, initial impairment magnitude, and their interaction predict the category effect. While all model variables were continuous in the regression model, for purposes of plotting the interaction in the figures, models with 0.5k~2.5k epochs of relearning were treated as <u>less</u> relearning (short gap) whereas those with more than 2.5k epochs of relearning were treated as <u>more</u> relearning (long gap). Similarly, for plotting purposes models with 0.5 or fewer connections lesioned were treated as mild-moderate impairment, whereas those with more than 0.5 connections lesioned were treated as moderate impairment.

We reviewed all 29 HSVE cases cited in Capitani et al.[2] and identified examination gap data in 19 of these. The time gap was calculated as the difference between the date at the onset of the disease and the date of the 1st neuropsychological examination (in years). The overall impairment was calculated as the averaged error rate on task performance collapsing across animal and artifact categories. The category effect was calculated as the accuracy difference between artifact and animal categories. All but two cases were assessed with the picture naming task. The same regression analysis was conducted with category effect as the dependent variable and time, overall impairment, and time*overall impairment as predictors. For plotting purposes, time gaps less than or equal to 1 year were binned as <u>short</u> while gaps long than 1 year were binned as <u>long</u>. Overall impairment was binned into mild-moderate and moderate-severe using accuracy of 0.5 as a cut-off.

For non-HSVE patients, we reviewed all 32 cases cited in Capitani et al.[2] and identified examination gap data from 17 cases. All but two cases were assessed with picture naming task. The same analysis for HSVE patients was conducted.

Supplementary References

1. Martin, A. & Chao, L. L. Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* **11,** 194–201 (2001).
2. Capitani, E., Laiacona, M., Mahon, B. Z. & Caramazza, A. What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn. Neuropsychol.* **20,** 213 (2003).
3. 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).
4. Laiacona, M., Capitani, E. & Barbarotto, R. Semantic category dissociations: A longitudinal study of two cases. *Cortex* **33,** 441–461 (1997).
5. 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).
6. 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).
7. McClelland, J. L. & Rogers, T. T. The parallel distributed processing approach to semantic cognition. *Nat Rev Neurosci* **4,** 310–322 (2003).
8. 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,** 385–396 (2011).
9. Rogers, T. T. & McClelland, J. *Semantic cognition: A parallel distributed processing approach*. (MIT Press, 2004).
10. Rogers, T. T. *et al.* Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychol. Rev.* **111,** 205–235 (2004).
11. James, T. W. & Gauthier, I. Auditory and Action Semantic Features Activate Sensory-Specific Perceptual Brain Regions. *Curr. Biol.* **13,** 1792–1796 (2003).
12. Binder, J. R. *et al.* Human temporal lobe activation by speech and nonspeech sounds. *Cereb. cortex* **10,** 512–528 (2000).
13. Davis, M. H. & Johnsrude, I. S. Hierarchical processing in spoken language comprehension. *J. Neurosci.* **23,** 3423–3431 (2003).
14. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12,** 718–724 (2009).
15. Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. Voice-selective areas in human auditory cortex. *Nature* **403,** 309–312 (2000).
16. Hickok, G. & Poeppel, D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4,** 131–138 (2000).
17. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.* **15,** 20–25 (1992).
18. Haxby, J. V *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (80-.).* **293,** 2425 (2001).
19. Reddy, L. & Kanwisher, N. Coding of visual objects in the ventral stream. *Curr. Opin. Neurobiol.* **16,** 408–414 (2006).
20. Lau, E. F., Phillips, C. & Poeppel, D. A cortical network for semantics: (de)constructing

the N400. *Nat Rev Neurosci* **9,** 920–933 (2008).
21. Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex* **1,** 1–47 (1991).
22. Gentileschi, V., Sperber, S. & Spinnler, H. Crossmodal agnosia for faimiliar people as a consequence of right infero-polar temporal atrophy. *Cogn. Neuropsychol.* **18,** 439–463 (2001).
23. Magnié, M.-N., Ferreira, C. T., Giusiano, B. & Poncet, M. Category specificity in object agnosia: Preservation of sensorimotor experiences related to objects. *Neuropsychologia* **37,** 67–74 (1998).
24. Goodale, M. A., Milner, A. D., Jakobson, L. S. & Carey, D. P. A neurological dissociation between perceiving objects and grasping them. *Nature* **349,** 154–156 (1991).
25. Goodale, M. A. *et al.* The nature and limits of orientation and pattern processing supporting visuomotor control in a visual form agnosic. *J. Cogn. Neurosci.* **6,** 46–56 (1994).
26. Goldenberg, G. & Hagmann, S. Tool use and mechanical problem solving in apraxia. *Neuropsychologia* **36,** 581–589 (1998).
27. Buxbaum, L. J. & Kalénine, S. Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann. N. Y. Acad. Sci.* **1191,** 201–18 (2010).
28. Binkofski, F. & Buxbaum, L. J. Two action systems in the human brain. *Brain Lang.* **127,** 222–229 (2013).
29. Schwartz, M. F. *et al.* Naturalistic action production following right hemisphere stroke. *Neuropsychologia* **37,** 51–66 (1998).
30. Buxbaum, L. J., Schwartz, M. F. & Carew, T. G. The role of semantic memory in object use. *Cogn. Neuropsychol.* **14,** 219–254 (1997).
31. Buxbaum, L. J., Kyle, K., Grossman, M. & Coslett, H. B. Left inferior parietal representations for skilled hand-object interactions: Evidence from stroke and corticobasal degeneration. *Cortex* **43,** 411–423 (2007).
32. 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).
33. Mahon, B. Z. *et al.* Action-related properties shape object representations in the ventral stream. *Neuron* **55,** 507–520 (2007).
34. Creem-Regehr, S. H. & Lee, J. N. Neural representations of graspable objects: are tools special? *Cogn. Brain Res.* **22,** 457–469 (2005).
35. 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).
36. 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).
37. 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).
38. 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).
39. 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).
40. 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).
41. 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).
42. Humphreys, G. F. & Lambon Ralph, M. A. Fusion and Fission of Cognitive Functions in the Human Parietal Cortex. *Cereb. Cortex* **25,** 3547–3560 (2015).
43. Chen, L. & Rogers, T. T. Revisiting domain-general accounts of category specificity in mind and brain. *Wiley Interdiscip. Rev. Cogn. Sci.* **5,** 327–344 (2014).
44. 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).
45. Adlam, A. L. R. *et al.* Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? *Brain* **129,** 3066–3080 (2006).
46. Patterson, K. *et al.* 'Presemantic' cognition in semantic dementia: Six deficits in search of an explanation. *J. Cogn. Neurosci.* **18,** 169–183 (2006).
47. Lambon Ralph, M. A., Sage, K., Jones, R. W. & Mayberry, E. J. Coherent concepts are computed in the anterior temporal lobes. *Proc. Natl. Acad. Sci.* **107,** 2717–2722 (2010).
48. Nestor, P. J., Fryer, T. D. & Hodges, J. R. Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage* **30,** 1010–1020 (2006).
49. Acosta-Cabronero, J. *et al.* Atrophy, hypometabolism and white matter abnormalities in semantic dementia tell a coherent story. *Brain* **134,** 2025–2035 (2011).
50. Noppeney, U. *et al.* Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain A J. Neurol.* **130,** 1138–1147 (2007).
51. Lambon Ralph, M. A., Ehsan, S., Baker, G. A. & Rogers, T. T. Semantic memory is impaired in patients with unilateral anterior temporal lobe resection for temporal lobe epilepsy. *Brain* **135,** 242–258 (2012).
52. 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).
53. 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* **19,** 2767–96 (2009).
54. Embleton, K. V, Haroon, H. A., Morris, D. M., 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).
55. 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,** 2728–38 (2010).
56. 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).
57. Halgren, E., Baudena, P., Heit, G., Clarke, M. & Marinkovic, K. Spatio-temporal stages in face and word processing. 1. Depth recorded potentials in the human occipital and parietal lobes. *J. Physiol.* **88,** 1–50 (1994).

58. Abel, T. J. *et al.* Direct physiologic evidence of a heteromodal convergence region for proper naming in human left anterior temporal lobe. *J. Neurosci.* **35,** 1513–1520 (2015).
59. Pobric, G., Jefferies, E. & Lambon Ralph, M. A. Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr. Biol.* **20,** 964–8 (2010).
60. 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,** 3802–17 (2015).
61. 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. U. S. A.* **104,** 20137–20141 (2007).
62. Saleem, K. S., Suzuki, W., Tanaka, K. & Hashikawa, T. Connections between Anterior Inferotemporal Cortex and Superior Temporal Sulcus Regions in the Macaque Monkey. **20,** 5083–5101 (2000).
63. Wakana, S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C. M. & Mori, S. Fiber Tract–based Atlas of Human White Matter Anatomy1. *Radiology* **230,** 77–87 (2004).
64. 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).
65. Seltzer, B. & Pandya, D. N. Parietal , Temporal , and Occipital Projections to Cortex of the Superior Temporal Sulcus in the Rhesus Monkey : A Retrograde Tracer Study. **463,** 445–463 (1994).
66. Schmahmann, J. D. & Pandya, D. *Fiber pathways of the brain*. (Oxford University Press, 2009).
67. 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,** 141–51 (2015).
68. Garcea, F. E. & Mahon, B. Z. Parcellation of left parietal tool representations by functional connectivity. *Neuropsychologia* **60,** 131–143 (2014).
69. Almeida, J., Fintzi, A. R. & Mahon, B. Z. Tool manipulation knowledge is retrieved by way of the ventral visual object processing pathway. *Cortex* **49,** 2334–2344 (2013).
70. Jones, M. N., Willits, J. & Dennis, S. Models of Semantic Memory. *Oxford Handb. Math. Comput. Psychol.* 232–254 (2015).
71. 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).
72. Warrington, E. K. & Shallice, T. Category specific semantic impairments. *Brain* **107,** 829–854 (1984).
73. Warrington, E. K. & McCarthy, R. Categories of knowledge.Further fractionations and an attempted integration. *Brain* **110,** 1273–1296 (1987).
74. Warrington, E. K. & McCarthy, R. Category-specific access dysphasia. *Brain* **106,** 859–878 (1983).
75. Rogers, T. T. *et al.* The structure and deterioration of semantic memory: a computational and neuropsychological investigation. *Psychol. Rev.* **111,** 205–235 (2004).
76. Rogers, T. T. & McClelland, J. L. Précis of semantic cognition: A parallel distributed processing approach. *Behav. Brain Sci.* **31,** 689–714 (2008).

77. McClelland, J. L. *et al.* Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* **14,** 348–356 (2010).
78. Devlin, J. T., Gonnerman, L. M., Andersen, E. S. & Seidenberg, M. S. Category-specific semantic deficits in focal and widespread brain damage: a computational account. *J. Cogn. Neurosci.* **10,** 77–94 (1998).
79. Hodges, J. R., Patterson, K., Oxbury, S. & Funnell, E. Semantic dementia progressive fluent aphasia with temporal lobe atrophy. *Brain* **115,** 1783–1806 (1992).
80. Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J. & Hodges, J. R. No right to speak? The relationship between object naming and semantic impairment: Neuropsychological evidence and a computational model. *J. Cogn. Neurosci.* **13,** 341–356 (2001).
81. Plaut, D. C. Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cogn. Neuropsychol.* **19,** 603–639 (2002).
82. Hinton, G. E. Learning distributed representations of concepts. in *Proceedings of the eighth annual conference of the cognitive science society* **1,** 12 (Amherst, MA, 1986).
83. Rumelhart, D. E. Brain style computation: Learning and generalization. in *An introduction to neural and electronic networks* 405–420 (Academic Press Professional, Inc., 1990).
84. Rumelhart, D. E. & Todd, P. M. Learning and connectionist representations. *Atten. Perform. XIV Synerg. Exp. Psychol. Artif. Intell. Cogn. Neurosci.* 3–30 (1993).
85. Rogers, T. T. & McClelland, J. L. A parallel distributed processing approach to semantic cognition: Applications to conceptual development. *Build. object Categ. Dev. time* 335–387 (2005).
86. Barsalou, L. W., Simmons, W. K., Barbey, A. K. & Wilson, C. D. Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* **7,** 84–91 (2003).
87. Barsalou, L. W. Grounded Cognition. *Annu. Rev. Psychol.* **59,** 617–645 (2008).
88. Ralph, M. A. L. Neurocognitive insights on conceptual knowledge and its breakdown. *Philos. Trans. R. Soc. London B Biol. Sci.* **369,** 20120392 (2014).
89. Lambon Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.*
90. Garcea, F. E. & Mahon, B. Z. Parcellation of left parietal tool representations by functional connectivity. *Neuropsychologia* **60,** 131–143 (2014).
91. 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).
92. 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).
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,** 1583–97 (2010).
94. 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).
95. 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).
96. Roberts, D. Exploring the link between visual impairment and pure alexia. (University of

Manchester, 2009).
97. Humphreys, G. W. & Forde, E. M. E. Category specificity in mind and brain? *Behav. Brain Sci.* **24,** 497–509 (2001).
98. Arguin, M., Bub, D. & Dudek, G. Shape Integration for Visual Object Recognition and Its Implication in Category-Specific Visual Agnosia. *Vis. cogn.* **3,** 221–276 (1996).
99. Viggiano, M. P., Costantini, A., Vannucci, M. & Righi, S. Hemispheric asymmetry for spatially filtered stimuli belonging to different semantic categories. *Cogn. Brain Res.* **20,** 519–524 (2004).
100. Mahon, B. Z. & Caramazza, A. Concepts and categories: A cognitive neuropsychological perspective. *Annu. Rev. Psychol.* **60,** 27–51 (2009).
101. Martin, A. The representation of object concepts in the brain. *Annu. Rev. Psychol.* **58,** 25–45 (2007).
102. Gerlach, C. A review of functional imaging studies on category specificity. *J. Cogn. Neurosci.* **19,** 296–314 (2007).
103. Joseph, J. E. Functional neuroimaging studies of category specificity in object recognition: a critical review and meta-analysis. *Cogn. Affect. Behav. Neurosci.* **1,** 119–136 (2001).
104. Brett, M., Christoff, K., Cusack, R. & Lancaster, J. Using the Talairach atlas with the MNI template. *Neuroimage* **13,** 85 (2001).
105. 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).
106. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38,** 95–113 (2007).
107. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Learning representations by back-propagating errors*. (MIT Press, Cambridge, MA, USA, 1988).