Why Bilateral Damage Is Worse than Unilateral Damage to the Brain

Anna C. Schapiro[1,2], James L. McClelland[1], Stephen R. Welbourne[3], Timothy T. Rogers[4], and Matthew A. Lambon Ralph[3]

Abstract

■ Human and animal lesion studies have shown that behavior can be catastrophically impaired after bilateral lesions but that unilateral damage often produces little or no effect, even controlling for lesion extent. This pattern is found across many different sensory, motor, and memory domains. Despite these findings, there has been no systematic, computational explanation. We found that the same striking difference between unilateral and bilateral damage emerged in a distributed, recurrent attractor neural network. The difference persists in simple feedforward networks, where it can be understood in explicit quantitative terms. In essence, damage both distorts and reduces the magnitude of relevant activity in each hemisphere. Unilateral damage reduces the relative magnitude of the contribution to performance of the damaged side, allowing the intact side to dominate performance. In contrast, balanced bilateral damage distorts representations on both sides, which contribute equally, resulting in degraded performance. The model's ability to account for relevant patient data suggests that mechanisms similar to those in the model may operate in the brain. ■

INTRODUCTION

There are indications from decades of research and clinical study that a unilateral lesion to a brain area can produce a very subtle—sometimes clinically insignificant—neurocognitive impairment, whereas a bilateral lesion may generate a profound deficit. Perhaps the best known case example arose in the 1950s, when Scoville produced a profound memory loss in patient HM by removing the hippocampus bilaterally. Scoville's motivation for publishing these findings was not only to warn other neurosurgeons of these clinically dramatic consequences but also to note that they ran counter to the expectations arising from patients after unilateral resections, in whom the observed memory deficit is relatively mild (Scoville & Milner, 1953; Milner & Penfield, 1955).

Why are there such different consequences of unilateral and bilateral lesions? One possibility is that a bilateral neural system is simply redundant enough across hemispheres to accommodate the amount of damage corresponding to complete unilateral removal. Although this may be part of the story, we will review findings strongly suggesting that bilateral lesions are much more detrimental than unilateral lesions even when unilateral lesions involve the same total amount of damage as bilateral lesions. This leads us to consider the possibility that there is something special about unilateral relative to bilateral damage. The computational exploration reported here provides evidence for this possibility, demonstrating that bilateral damage is much more detrimental than unilateral damage in a range of distributed neural network models even when the amount of damage is equated.

To anchor the investigation with a concrete, nontrivial test case, we applied the computational work to the domain of semantic memory. This choice was motivated by the availability of a considerable body of relevant patient and neuroscience data. In addition, the new computational investigation built on earlier modeling work in which damage to a single pool of hidden units was used to simulate effects of bilateral anterior temporal lobe (ATL) atrophy in semantic dementia (SD; Rogers et al., 2004). The findings we obtain generalize across tasks and to a range of different network architectures. Our results may help, therefore, to explain the different effects of unilateral and bilateral damage that have been reported in many human and animal studies across a range of task domains and brain areas. Such differences apply not only to effects of damage to the human hippocampus (Scoville & Milner, 1957) and ATL as examined in detail here but also to effects of damage in, for example, rat hippocampus (Li, Matsumoto, & Watanabe, 1999), primate anterior temporal cortex (Klüver & Bucy, 1939), primate frontal lobe (Warden, Barrera, & Galt, 1942), primate auditory cortex (Heffner & Heffner, 1986), human frontal lobe (D'Esposito, Cooney, Gazzaley, Gibbs, & Postle, 2006), and human amygdala (Adolphs & Tranel, 2004).

[1]Stanford University, [2]Princeton University, [3]University of Manchester, [4]University of Wisconsin-Madison

© 2013 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 25:12, pp. 2107–2123
doi:10.1162/jocn_a_00441

We note, in advance, that our results will apply to effects of unilateral versus bilateral damage to the extent that there is a balance of involvement of brain structures across the hemispheres. In cases where a function is completely lateralized, damage to the hemisphere that specializes in the function will naturally be more disruptive to performance than balanced bilateral damage. But, as we shall show, a degree of asymmetry can still be consistent with a relatively milder deficit after unilateral damage, even to the more dominant hemisphere. We will demonstrate this by examining the effect of unilateral versus bilateral damage on two semantic tasks. Performance in one of these tasks (word-to-picture matching) appears to be subserved by the ATL approximately symmetrically, whereas performance in the other (naming) is underpinned to a greater degree by the left ATL (Lambon Ralph, Ehsan, Baker, & Rogers, 2012; Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001; Seidenberg et al., 1998). This is captured in the model before lesioning through a mild structural asymmetry.

The Role of Anterior Temporal Regions in Semantic Representation

As noted above, we chose the domain of semantic memory as a test case for the simulations and, in particular, the role of left and right ATL regions in supporting semantic representations. The degradation of semantic memory in SD and herpes simplex virus encephalitis is associated with bilateral damage to and hypometabolism of the ATLs (Mion et al., 2010; Rohrer et al., 2009; Noppeney et al., 2007; Nestor, Fryer, & Hodges, 2006). Behavioral data from these patients have suggested a model in which concepts are formed through the convergence of sensory, motor, and verbal experience in a transmodal representational hub in the ATL (Rogers et al., 2004).

Although previously overlooked, there is now a growing consensus that this transmodal ATL hub contributes critically to semantic cognition (Patterson, Nestor, & Rogers, 2007). This emerging view reflects a convergence of the established clinical data on SD, herpes simplex virus encephalitis, etc., with contemporary neuroscience studies. Data from TMS and functional neuroimaging indicate that both left and right ATL regions contribute to semantic memory. For example, the multimodal, selective semantic impairment of SD can be mimicked in neurologically intact participants by applying rTMS to the left or right lateral ATL (Pobric, Jefferies, & Lambon Ralph, 2007, 2010; Lambon Ralph, Pobric, & Jefferies, 2009). Likewise, when using techniques that avoid (e.g., PET or MEG) or correct for the various methodological issues associated with successful imaging of the ATL (Visser, Jefferies, & Lambon Ralph, 2010; Devlin et al., 2000), studies find bilateral ATL activation for semantic processing across multiple verbal and nonverbal modalities (Visser & Lambon Ralph, 2011; Binney, Embleton, Jefferies, Parker, & Lambon Ralph, 2010; Visser, Embleton, Jefferies, Parker, & Lambon Ralph, 2010; Sharp, Scott, & Wise, 2004; Marinkovic et al., 2003; Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996).

This domain is a highly pertinent one for the current investigation given that unilateral and bilateral ATL damage have very different consequences on semantic performance (mirroring the pattern found in other cognitive domains). This conundrum is summarized in Figure 1. In terms of clinical presentation, the bilateral atrophy of SD (Figure 1A) leads to a notable, selective semantic impairment (Patterson & Hodges, 1992; Snowden, Goulding, & Neary, 1989; Warrington, 1975). In contrast, patients with unilateral ATL damage (in this example, Figure 1B, a full en bloc ATL resection for intractable temporal lobe epilepsy [TLE]) do not report comprehension impairment (Hermann, Seidenberg, Haltiner, & Wyler, 1995; Hermann et al., 1994) but do complain of amnesia and anomia, especially following left ATL resection (Glosser, Salvucci, & Chiaravalloti, 2003; Martin et al., 1998; Seidenberg et al., 1998). The two contrastive patients in Figure 1 are particularly pertinent given that the unilateral case represents a complete removal of ATL tissue (Figure 1B) but had little or no drop in accuracy on standard semantic tests (see below), whereas the bilateral case represents only partial tissue loss (Figure 1A) but exhibited considerable semantic impairment. Although an exact comparison of the extent of relevant tissue loss is not possible, the notable difference in the extent of deficit despite comparable lesion size suggests that size is only part of the story.

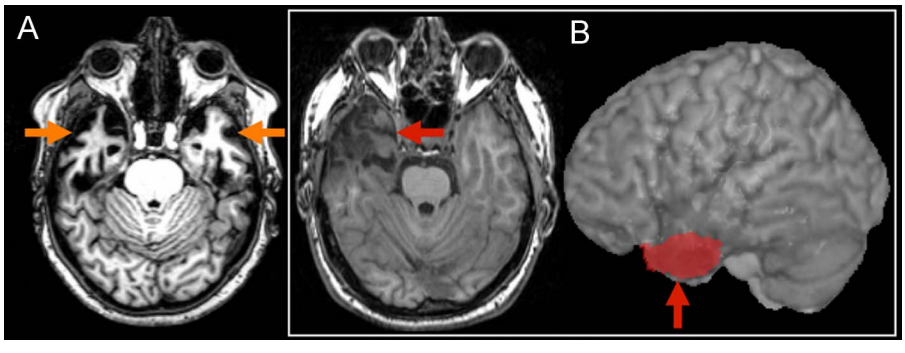


Figure 1. Example scans of patients with unilateral and bilateral ATL damage. (A) An example axial MRI for a patient with semantic impairment resulting from bilateral ATL atrophy (orange arrows). (B) A comparable axial slice from a patient following ATL unilateral resection for TLE (red arrow). The red region on the lateral view shows the resected area. Adapted from Figure 1 of Lambon Ralph et al. (2012).

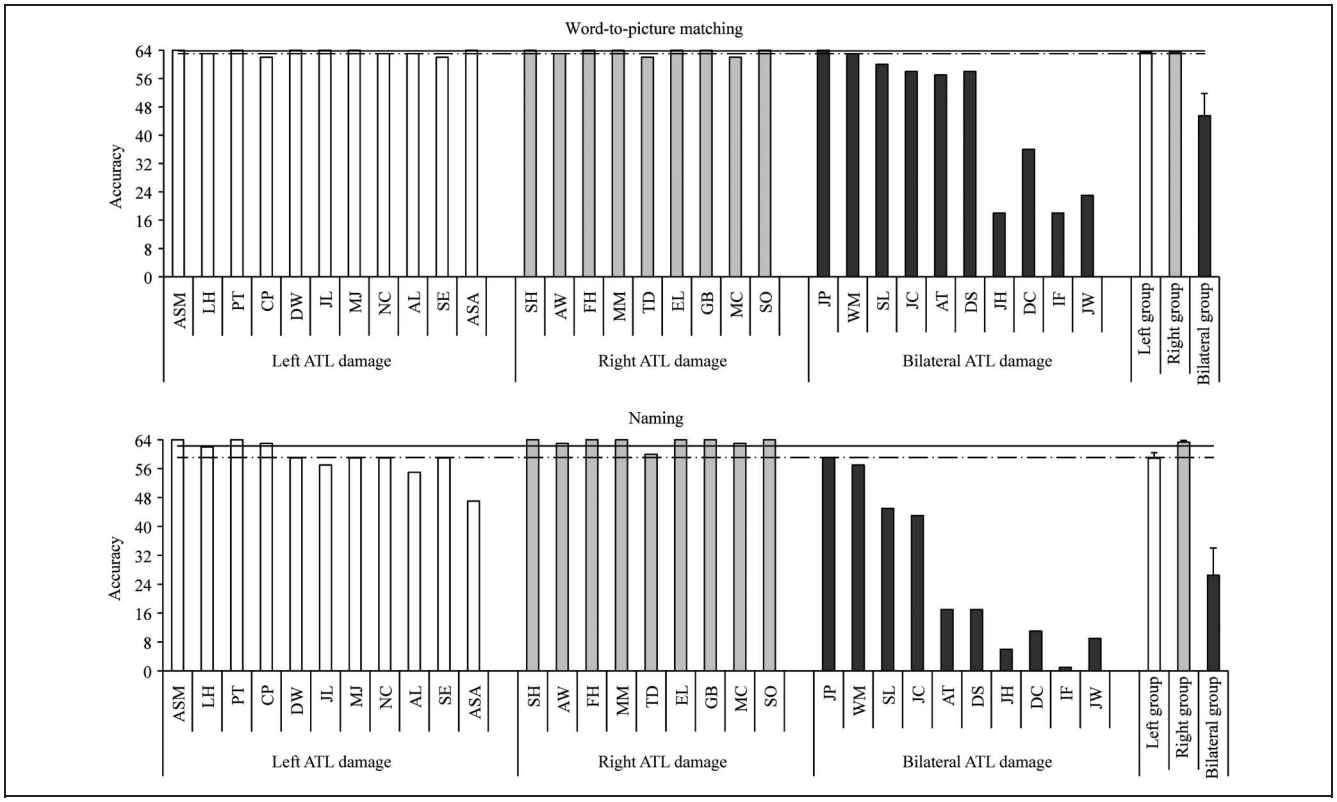


Figure 2. Patient performance. The proportion of items correct (out of 64) on naming and word-to-picture matching tasks of patients with unilateral left, unilateral right, and bilateral ATL damage. Averages for each patient group are included on the far right of each graph, with error bars indicating ±1 *SEM*. The full horizontal line denotes control–participant mean performance, and the dashed line shows 2 *SD* below the control mean. Adapted from Figure 2 of Lambon Ralph et al. (2010).

These clinical observations are striking but need to be treated with some caution given that clinical assessment of TLE patients tends to focus on naming and episodic memory and rarely on comprehension (Giovagnoli, Erbetta, Villani, & Avanzini, 2005). Two recent case series neuropsychological studies, therefore, assessed this clinically apparent contrast via in-depth explorations of semantic function. The results not only supported the clinical observations but also provided a bridge to the neuroimaging and rTMS data noted above. The first study (Lambon Ralph, Cipolotti, Manes, & Patterson, 2010) investigated patients' accuracy on a range of tasks taken from a multimodal semantic battery commonly used to assess patients with SD (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000). A case series of patients with unilateral left or right ATL damage (of mixed etiology) performed at or very close to normal levels of accuracy on the various comprehension tasks, whereas most SD patients were impaired (Figure 2). In addition, like SD patients with asymmetrically left-biased ATL atrophy (Lambon Ralph et al., 2001), the left unilaterally damaged patients exhibited significantly greater anomia than their right-sided counterparts. The second study focused specifically on patients with left- or right-sided ATL resections for the treatment of TLE (Lambon Ralph et al., 2012). Again, on standard accuracy-based measures of semantic function, the vast majority of the resected TLE patients performed within the normal range. However, mild semantic dysfunction was revealed when either the semantic demands of the tasks were increased (by probing specific-level concepts, low frequency, or abstract items) or by measuring RTs as well as accuracy (the patients were two to three times slower than older controls even on the simpler semantic tasks). Performance on these more demanding semantic measures correlated significantly with the volume of resected tissue and, again, the left-resected patients exhibited significantly greater anomia than the right-sided cases. Together, these results provided us with a series of target phenomena to be captured in the computational model:

i. regions within the left and right ATL jointly support pan-modal semantic function in neurologically intact participants;
ii. unilateral resection or stimulation via rTMS generates detectable but mild semantic impairment;
iii. bilateral lesions generate significantly greater semantic impairment than unilateral damage, even when overall amount of damage is similar;
iv. left ATL damage generates greater naming impairment than equivalent levels of right ATL damage.

To fully account for the range of lesion scenarios, we also explored the possibility that, in some patients,

plasticity-related recovery after or during the course of brain damage may have affected postlesion performance (Keidel, Welbourne, & Lambon Ralph, 2010; Welbourne & Lambon Ralph, 2005; Wilkins & Moscovitch, 1978). Patients continue to engage in semantic processing after an acute lesion or, in degenerative cases, as their disease progresses, and this engagement may help to ameliorate the effects of damage (Welbourne & Lambon Ralph, 2005). We considered the effects of plasticity-related recovery in the model to demonstrate its applicability to these scenarios.

Overview of the Model

The model we used to account for the stark contrast between the behavior of patients with unilateral and bilateral damage is an extension of a model used previously to explain the deterioration of semantic task performance in SD (Rogers et al., 2004). It consisted of processing units that were organized into layers and connected as shown in Figure 3. The connection weights between the units were established by training the network with patterns derived from descriptions and visual properties of familiar objects. The input to the network came from the verbal descriptor and visual feature layers, which are considered analogous to cortical areas that subserve high-level verbal and visual information, respectively. The verbal descriptors were divided into sublayers that represent names, verbal descriptions of perceptual features, functional descriptors, and encyclopedic information.

The model extended the previous simulation architecture (Rogers et al., 2004) by dividing what was previously a single pool of hidden units into two: left demi-hub and right demi-hub. Units in these pools had no prespecified representations—rather, the pools developed distributed representations over the course of training that allowed the network to map between the features of the objects in the environment and to capture the similarity relations of the objects to one another. The layers of the model were highly interconnected, with complete bidirectional connectivity (all the units in one layer connected bidirectionally to all the units in another layer) between verbal descriptors and left demi-hub, visual features and left demi-hub, and visual features and right demi-hub. The projection from the verbal descriptors to the right demi-hub was complete, but the return projection had about 80% of all the possible connections. The difference between density of projection from the left demi-hub to verbal descriptors and that from the right demi-hub to verbal descriptors caused the left demi-hub, over the course of training, to play a somewhat more important role in tasks like naming from visual input that require a verbal response, aligning this model with our previous one on laterality effects in SD (Lambon Ralph et al., 2001).

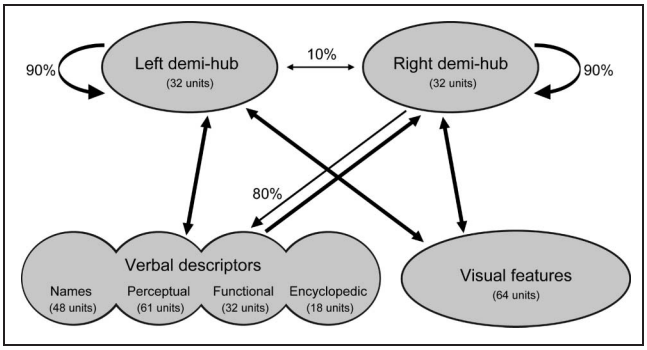


Figure 3. Model architecture. The model consists of two input/output layers, verbal descriptors and visual features, and two hidden layers, left demi-hub and right demi-hub. Arrows indicate full connectivity from all the units in the projecting to the receiving layer, except where there are percentages written which specify a different degree of connectivity.

An important feature of the model was the stipulation that the density of the recurrent connections within each hidden layer was 90% whereas the density of the projection between the hidden layers was only 10%. This stipulation is consistent with the fact that there are more intrahemispheric than interhemispheric connections in cortex (Gee et al., 2011; Lewis, Theilmann, Sereno, & Townsend, 2009; Braitenberg & Schüz, 1998). As we shall see, our model suggests that this difference may be an important part of the reason why bilateral damage is more disruptive than unilateral in neural populations with recurrent connections.

In what follows, we first demonstrate that our model simulates an advantage of unilateral compared with bilateral damage when overall lesion severity is equated. We proceed to consider this phenomenon in a range of network architecture variants, one of which is sufficiently simple to allow us to express mathematically the fundamental basis for the advantage of unilateral lesions. Our analysis shows why the advantage occurs in a wide range of neural network architectures and, therefore, why it is natural to expect that this advantage should occur in real biological neural networks such as those found in the brains of humans and other animals.

METHODS

The reported simulations were conducted using the LENS neural network simulation environment (www.stanford.edu/group/mbc/LENSManual/index.html). A network very similar to the one used by Rogers et al. (2004) was reimplemented within this environment. The following describes the main simulation reported first in the Results section. Other simulations used similar methods; where details differ from the main simulation, this is stated in the Results section.

Network Initialization

In setting up the network, complete projections were created by simply connecting each unit in the projecting

layer to each unit in the receiving layer. For sparser projections, for example, the 10% connection between the two hidden layers, each unit in one hidden layer was connected to each unit in the other hidden layer with an independent probability of 0.1. As a result, each instantiation of a network had a different pattern of connectivity and a somewhat variable proportion of connections. Weights were also assigned randomly before training, with values chosen uniformly between −1 and 1.

Activation Dynamics

Each unit in the verbal descriptors and visual features layers represented a unique verbal or visual property of an object. To present an object to the network, the units representing the features of the object took on the value 1, and otherwise had the value 0. The presentation of an item to the network was divided into seven time intervals; in each interval, activation of each unit i was calculated by first determining the net input, n_i, to the unit:

$$n_i = \sum_j a_j w_{ij}$$

where j indexes all units that project to unit i, a_j is the activation of unit j, and w_{ij} is the weight of the connection from unit j to unit i. The activation of unit i was then calculated based on the following logistic function, which bounds activation between 0 and 1:

$$\alpha_i = \frac{1}{1 + e^{-n_i}}.$$

All units were assigned a fixed, untrainable bias of −2, which deducts 2 from each unit's net input. This bias parameter, in the absence of other input, keeps unit activations on the low end of their range (at approximately 0.12).

Training

The network was trained and tested using 48 patterns organized into six natural taxonomic categories (birds, mammals, vehicles, household objects, tools, fruits). Patterns were generated from the visual feature and verbal descriptor prototypes used by Rogers et al. (2004), which were designed to reflect the similarity relations of typical objects in the six categories. The similarity of patterns within categories varies across categories as it does for real objects in the corresponding categories. Patterns were constructed exactly as in Rogers et al. (2004) with the exception that object names were not given at different levels of specificity in the present work—each of the 48 objects was assigned its own unique name.

Training the network involved teaching it to "bring to mind" the full information associated with an item, either from the name, the visual pattern, or from non-name verbal descriptors of the item. Accordingly, in each training trial, the units corresponding to one of these three inputs (units in the name, other verbal descriptors, or visual features layer) were set (hard-clamped) to the specified input values for three intervals of processing. The clamps were then removed, and the network cycled for four more intervals, so that all unit activations were then determined by the propagation of activation through the network's connections. During the last two intervals of the seven total, the target values for all visual and verbal patterns (including name) were applied, and error derivatives for weights in the network were calculated. Training was conducted using the standard back propagation gradient descent learning algorithm for recurrent networks in LENS, with 200,000 sweeps through each of the 48 training patterns presented in a random order (each sweep is called an epoch). Weights were updated at the end of each epoch, except during recovery, when a finer grain of detail is necessary and weights were updated after every pattern. The model was trained with a learning rate of 0.005, weight decay of 0.0002 per epoch during training (0.000001 per pattern during recovery), and no momentum. Zero-mean 0.5 standard deviation Gaussian noise was added to the net input of each unit at each activation update. Gaussian noise with the same standard deviation was also included during testing unless otherwise specified.

Testing

Testing for picture naming assessed the network's ability to activate the correct name unit from visual input. This was implemented by clamping the visual feature units to the values corresponding to a trained object, allowing activation to propagate for three intervals, then removing the input and allowing activation to continue to propagate for four more intervals. The unit with the highest activation was treated as the model's name response to the presented picture.

The test for word-to-picture matching compared the pattern across the hidden layers produced by an item's name with the patterns produced by several alternative visual input patterns. A trial was scored correct if the pattern produced by the name for an item was more similar to the pattern produced by the visual feature input pattern for the same item than it was to the pattern produced by the visual input pattern for seven distractor items. To accomplish this, the target name unit was clamped for the first three processing intervals; input was then removed, and activation continued to propagate for four intervals. The resulting pattern of activation across all units in both hidden layers was recorded. This process was repeated with visual feature input for the target concept and each of the seven remaining items in the same semantic category (acting as the foils in the assessment). The resulting patterns of hidden unit activation can be thought of as nine vectors (one resulting from presentation of the

name and eight from presentation of the visual features), where each unit's activation is an element of the vector. The cosine between the vector from the name input and the vector from each of the visual feature inputs was used as the measure of similarity.

The cosine indicates the degree of similarity between two vectors, independent of their magnitudes, and is defined as the dot product of the two vectors divided by the product of their magnitudes:

$$\cos(V_1, V_2) = \frac{V_1 \cdot V_2}{||V_1||||V_2||}$$

where $V_1 \cdot V_2 = \sum_i V_{1i} V_{2i}$ and $||V|| = \sqrt{\sum_i V_i^2}$. The set of visual features that resulted in the highest cosine with the name was interpreted as the model's choice of picture.

Simulation of Behavioral Deficits under Damage

To investigate the effects of damage on task performance, the model was damaged at several levels of severity, by permanently removing connections into and out of a layer with a specified probability. Damage of, for example, 50% to a hidden layer consisted of removing connections into and out of each unit in that layer with a probability of 0.5. For progressive damage simulations, where lesions were applied incrementally and the network was tested as damage progressed, damage levels refer to the total cumulative proportion of connections removed up to that point. Bias weights were not lesioned, although results were the same when lesions to bias weights were allowed.

In one of the simulations reported below, damage was simulated by the removal of whole units rather than individual connections. In this case, damage to a layer consisted of unit removal with a specified probability. As with other forms of damage, removal was probabilistic, such that the actual number of units removed could vary.

Damage was applied to one or both of the hidden (demi-hub) layers. For a given severity level, a bilateral lesion was produced by applying this probability uniformly across the two hidden layers, whereas unilateral left and right lesions were produced by applying damage with twice this probability to the connections into and out of only one of the two hidden layers. Following this procedure, the largest possible unilateral lesion corresponded to a 50% bilateral lesion. However, we extended bilateral lesions beyond this severity level to 100% to simulate more severe stages of SD and other bilateral diseases. In the case of bilateral damage, the connections between the two hidden layers were removed at the associated unilateral damage level to ensure that the total number of connections was approximately equal in the unilateral and bilateral damage cases. For example, if 50% of the connections between hidden layers were removed for a unilateral lesion, 50% of those connections were again removed in the corresponding case of bilateral damage (as opposed to 25% removal twice—one for each hidden layer—which would result in only 44% of these connections being lesioned).

To examine how experience after damage or during the course of degeneration might affect performance in the model, we conducted simulations using varying amounts of recovery after performing a simulated lesion. During this progressive damage, retraining occurred after each round of lesioning, and the network was tested immediately before and after an episode of damage and the results of the two tests were averaged. This was done to approximate the more continuous mix of damage and exposure to the environment that patients with neurodegenerative disease experience.

Results reflect the performance of 10 trained networks initialized with different random seeds. Each network was tested after each combination of lesion type (left, right, or bilateral), severity, and recovery amount (where these were varied), and the results were averaged over the 10 separate networks.

RESULTS

Differential Effects of Unilateral versus Bilateral Damage

Our first simulation demonstrates a striking difference between the effects of unilateral and bilateral lesions in the recurrent network architecture shown in Figure 3. Figure 4 summarizes simulated picture naming and word-to-picture matching performance as a function of different levels of unilateral left, unilateral right, and bilateral connection damage following training of the recurrent network. Results are presented separately for different amounts of recovery after each simulated lesion. Networks with bilateral damage performed much worse than those with unilateral damage on both tasks, even with the same total amount of damage and recovery (see Table 1 for statistics). Bilateral damage caused performance to fall relatively quickly, whereas even a complete unilateral lesion, especially with some recovery, had a relatively small effect. The difference between the effect of unilateral and bilateral damage became more pronounced with increasing lesion severity.

One additional feature of these results is that in none of these cases did performance drop off immediately as soon as there was any damage. The use of distributed representations in the model, encouraged by the presence of noise and weight decay in training and by retraining after damage, made the model quite robust. This graceful degradation (Farah & McClelland, 1991; Hinton & Sejnowski, 1983) is clear in patients as well—SD patients demonstrate a drop in semantic performance over time and have some remaining semantic ability even after extensive atrophy.

There was also a difference, especially in naming, between left-sided and right-sided unilateral damage. Because of the somewhat greater role of the left demi-hub in mapping visual to verbal features, damage to the left

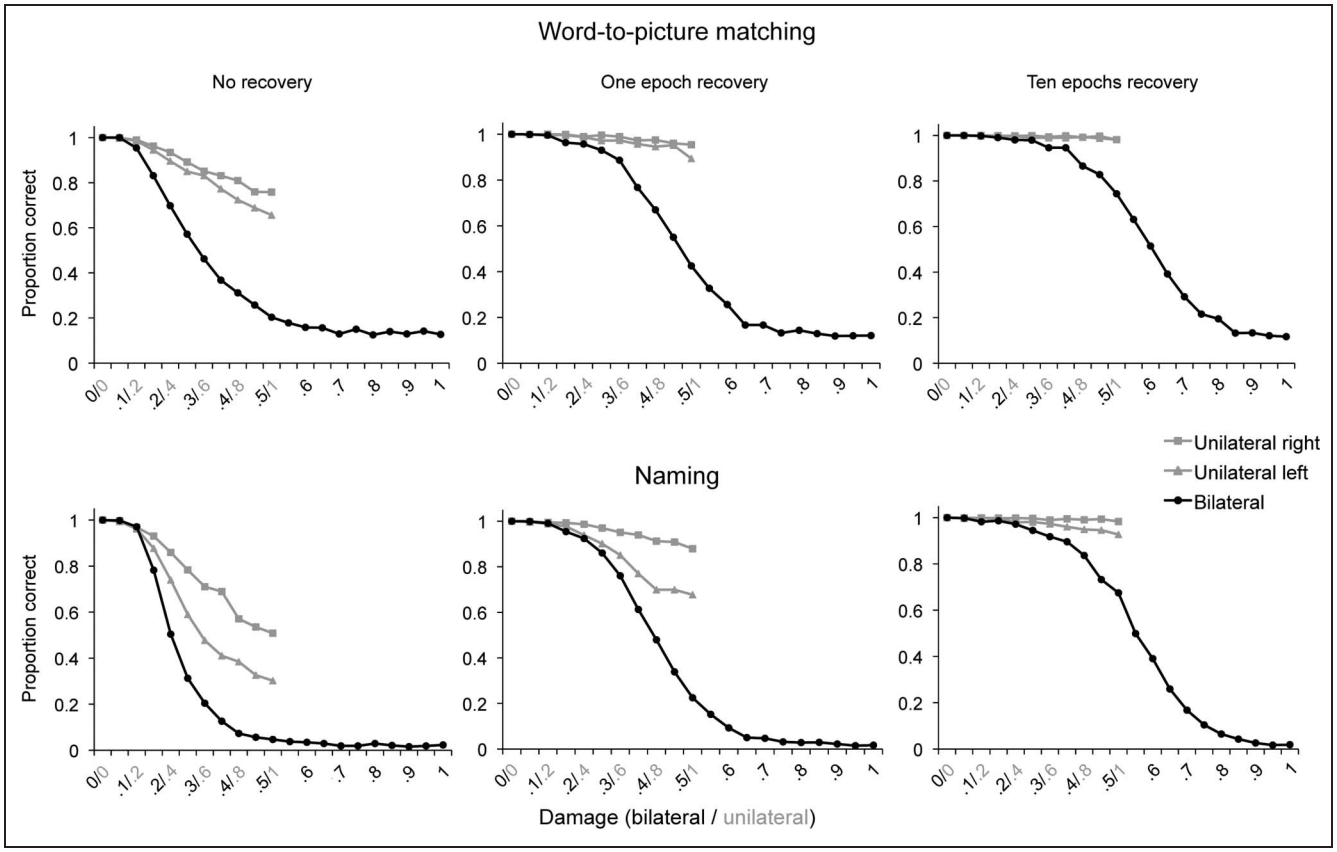


Figure 4. Performance of the network on word-to-picture matching and naming with unilateral and bilateral damage. Proportion correct on these tasks is shown for 20 levels of damage in each graph, labeled by the proportion of connections removed from units in both hidden layers or one hidden layer. Deterioration of performance is shown with no recovery throughout the course of damage, one epoch of recovery between each episode of damage, and 10 epochs between each episode of damage.

side of the network was more detrimental to naming than damage to the right side. There was also a small difference between left-sided and right-sided unilateral damage in word-to-picture matching.

The simulation results presented so far employed cumulative damage with retraining after each stage of lesioning. Some of the unilateral patients, for example, those with long-standing epilepsy followed by resection, had damage that was only partially progressive and some, for example, those with stroke, had damage that would not have been progressive at all. We ran additional simulations to test whether the model would perform differently if damage were applied all at once, followed by a period of recovery (to mimic an acute neurological event followed by spontaneous partial recovery) instead of interleaving damage and retraining. The results of the nonprogressive damage simulations were very similar to those for the progressive damage simulations (see Appendix 1).

Effect of Lesions on Similarity Structure

Why does a unilateral lesion have less effect on performance than the same severity of lesion distributed bilaterally? To understand these differential effects of unilateral and bilateral damage in the model, we examined how the similarity structure of its internal representations changed with damage. This is useful because distortions to the similarity structure were the source of errors in word-to-picture matching and contributed to errors in naming (for naming, weights from the hidden layers to the name layer also contribute). For this analysis, we used a network architecture in which the differential connectivity from the left and right demi-hubs to verbal output was eliminated for simplicity, and no recovery was allowed, although similar results were obtained with an architecture that retained the differential connectivity and for networks with various levels of recovery.

We obtained the pattern of activation across both demi-hubs for presentations of each object's name and each object's visual input pattern, as in the word-to-picture matching task, then calculated the cosine similarity of the pattern produced by each name with the pattern produced by each visual input pattern, averaging the results across 10 networks. This yielded a 48 × 48 matrix, with each item corresponding to a row of the matrix (visual feature input) and a column (name input), as shown in Figure 5. Objects in the same category are grouped together. High cosine values along the diagonal of the matrix for the intact network (left-most panel of

Table 1. Analysis of Variance for Figure 4 Results

	df	*F*	*p*
No Recovery			
Word-to-picture matching			
Damage	10, 90	301.2	<.0001
Bilateral vs. mean unilateral	1, 9	264.0	<.0001
Damage * unilateral vs. bilateral	10, 90	74.57	<.0001
Right vs. left unilateral	1, 9	4.53	.0622
Naming			
Damage	10, 90	888.7	<.0001
Bilateral vs. mean unilateral	1, 9	277.5	<.0001
Damage * unilateral vs. bilateral	10, 90	68.07	<.0001
Right vs. left unilateral	1, 9	62.52	<.0001
One Epoch Recovery			
Word-to-picture matching			
Damage	10, 90	405.9	<.0001
Bilateral vs. mean unilateral	1, 9	849.1	<.0001
Damage * unilateral vs. bilateral	10, 90	190.6	<.0001
Right vs. left unilateral	1, 9	16.49	.003
Naming			
Damage	10, 90	372.1	<.0001
Bilateral vs. mean unilateral	1, 9	692.2	<.0001
Damage * unilateral vs. bilateral	10, 90	196.0	<.0001
Right vs. left unilateral	1, 9	119.40	<.0001
Ten Epochs Recovery			
Word-to-picture matching			
Damage	10, 90	81.85	<.0001
Bilateral vs. mean unilateral	1, 9	257.8	<.0001
Damage * unilateral vs. bilateral	10, 90	52.81	<.0001
Right vs. left unilateral	1, 9	5.39	.045
Naming			
Damage	10, 90	172.3	<.0001
Bilateral vs. mean unilateral	1, 9	305.47	<.0001
Damage * unilateral vs. bilateral	10, 90	95.56	<.0001
Right vs. left unilateral	1, 9	73.45	<.0001

Two-factor repeated-measures ANOVAs treat network initializations as the random effects factor, assessing overall effect of damage extent, effect of unilateral (averaged across right and left damage) compared with bilateral damage, the interaction between damage extent and the difference between unilateral and bilateral damage, and the difference between right and left unilateral damage. Damage levels include zero to full unilateral damage over 11 increments (as in Figure 4 unilateral damage).

figure) indicated that the network was able to associate an object's visual features correctly with its name. Although the cosines are generally higher within than between category, the correct (diagonal) entry is strongest in each row and each column before damage. As damage increased, the similarity structure of the unilaterally lesioned network was fairly well preserved, allowing it to distinguish an item from the others in its category with fairly high accuracy, even after a complete lesion to one side of the network. The bilaterally damaged network, in contrast, lost the appropriate similarity structure rapidly; with a 50% lesion to each side of the network, the similarity structure seen with the intact network was nearly completely obliterated, accounting for its very poor performance.

Simpler Networks and Representation Analyses

The preceding analysis shows that bilateral damage degrades the similarity structure of the network's internal representations, whereas unilateral damages leaves this largely intact. What causes this phenomenon? The bidirectional and recurrent connections in the current model make it difficult to answer this question, so we next considered a simpler network architecture in which all connections were feedforward (i.e., projecting in one direction, from input toward output).

Feedforward Network

In the feedforward network, each of the layers that was previously used as both input and output was duplicated so that one copy could be used for the input and the other for the output (see Figure 6). In addition, there were no recurrent connections within or between the two hidden layers, no recovery was allowed, and the connections into and out of the two hidden layers were complete and symmetric. Although training affects both input and output connections to the hidden units in this model, only the input connections affect the internal representations, allowing a focused analysis of the fundamental basis for the unilateral versus bilateral differences in word-to-picture matching in this version of the model. We trained 10 networks initialized with different random seeds as in the main simulation, then applied unilateral and bilateral damage to this feedforward architecture by removing connections as in previous simulations. During training and testing, activations were calculated in a single feedforward pass, and noise was included in processing as per the main simulation. Even in this simplified network there was still an advantage for unilateral compared with bilateral damage, both in word-to-picture matching and naming (Appendix 2).

Representation Fidelity

We chose a measure which we call "fidelity" to quantify the loss of similarity structure in this network. Because

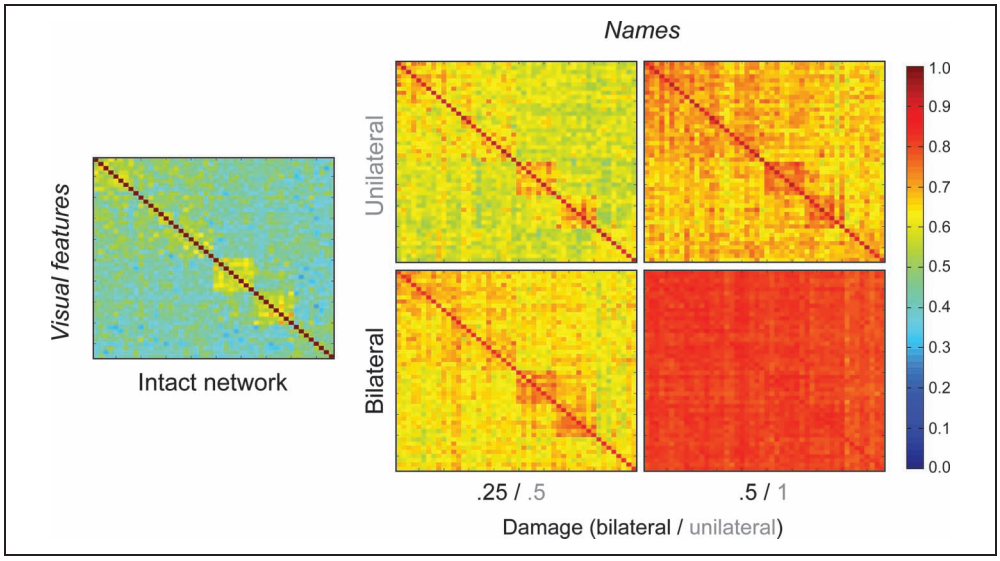


Figure 5. Degradation of similarity structure under unilateral and bilateral damage. The colors in the heatmaps represent the value of the cosines between hidden layer activation vectors after each set of visual features (indexed by rows) and each name (indexed by columns) were presented to the intact network and to the network after two levels of unilateral and bilateral damage. Names and visual features are clustered by category. Cosine values range from 0 (blue) to 1 (red).

the relative values of the on- and off-diagonal cosines in the matrices shown in Figure 5 govern the network's ability to perform semantic tasks like naming and word-to-picture matching, fidelity is defined as the difference between mean diagonal and off-diagonal cosine terms, including both within- and between-category off-diagonal entries (see Figure 7 caption for details). Although fidelity does not correspond directly to accuracy, greater fidelity is associated with more accurate performance. This measure is useful because it can be applied to the left and right demi-hubs separately to provide an assessment of the extent to which each side carries the appropriate similarity information on its own. Figure 7A shows the fidelity of an intact demi-hub on its own, a damaged demi-hub, the average of those two, and compares this to the average of bilaterally damaged demi-hubs (each of which produces fidelity similar to this average). We found that within a layer, fidelity was approximately equal to the fraction of connections remaining, with the result that the average fidelity was very similar in the unilateral and bilateral cases, equating for the total number of connections removed. (See Simplified Mathematical Analysis and Note 1 for an explanation of the divergence between average unilateral and bilateral fidelity at high levels of damage.)

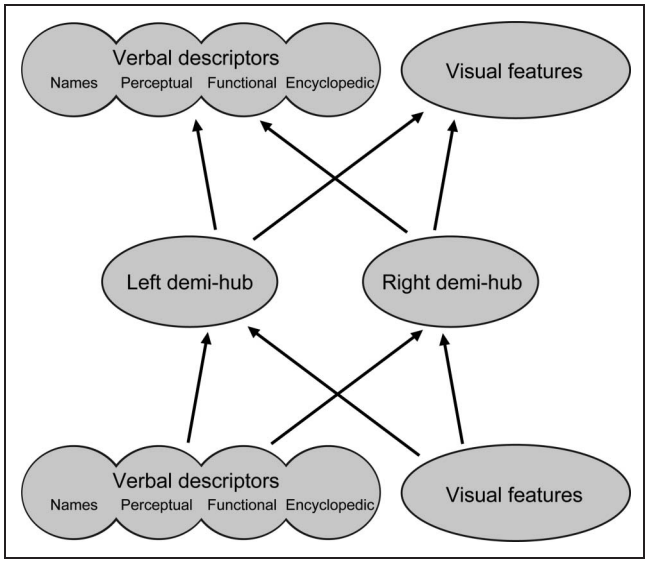


Figure 6. Feedforward model architecture.

If average fidelity across the left and right demi-hubs is as impaired after a unilateral lesion as it is after a bilateral lesion, why then is word-to-picture matching and naming more accurate after the unilateral lesion? The reason is that the fidelity calculated across both layers remains much higher in the unilateral case. Figure 7B shows the fidelity measure calculated across both layers of the feedforward network, either with unilateral damage or bilateral damage. As in word-to-picture matching and naming performance, we found that the fidelity was always higher in the unilateral than bilateral case (see Table 2). Interestingly, fidelity followed a U-shaped function, decreasing over low to moderate levels of damage, then increasing again as the extent of damage approached its maximal value. This occurs because variation from item to item in the pattern of activation of the units in the damaged layer diminishes as the fidelity decreases, and so these units contribute less and less to the overall fidelity of the representations. By the time their contributions become very weak, the intact side of the network completely dominates performance. Thus, a layer with a large amount of unilateral damage carries little information about the items but also contributes little to the overall performance of the network because it does not provide a differential response across items. This is the essence of the reason why unilateral damage is so much more benign than bilateral damage: With more and more damage to just one side of the net-

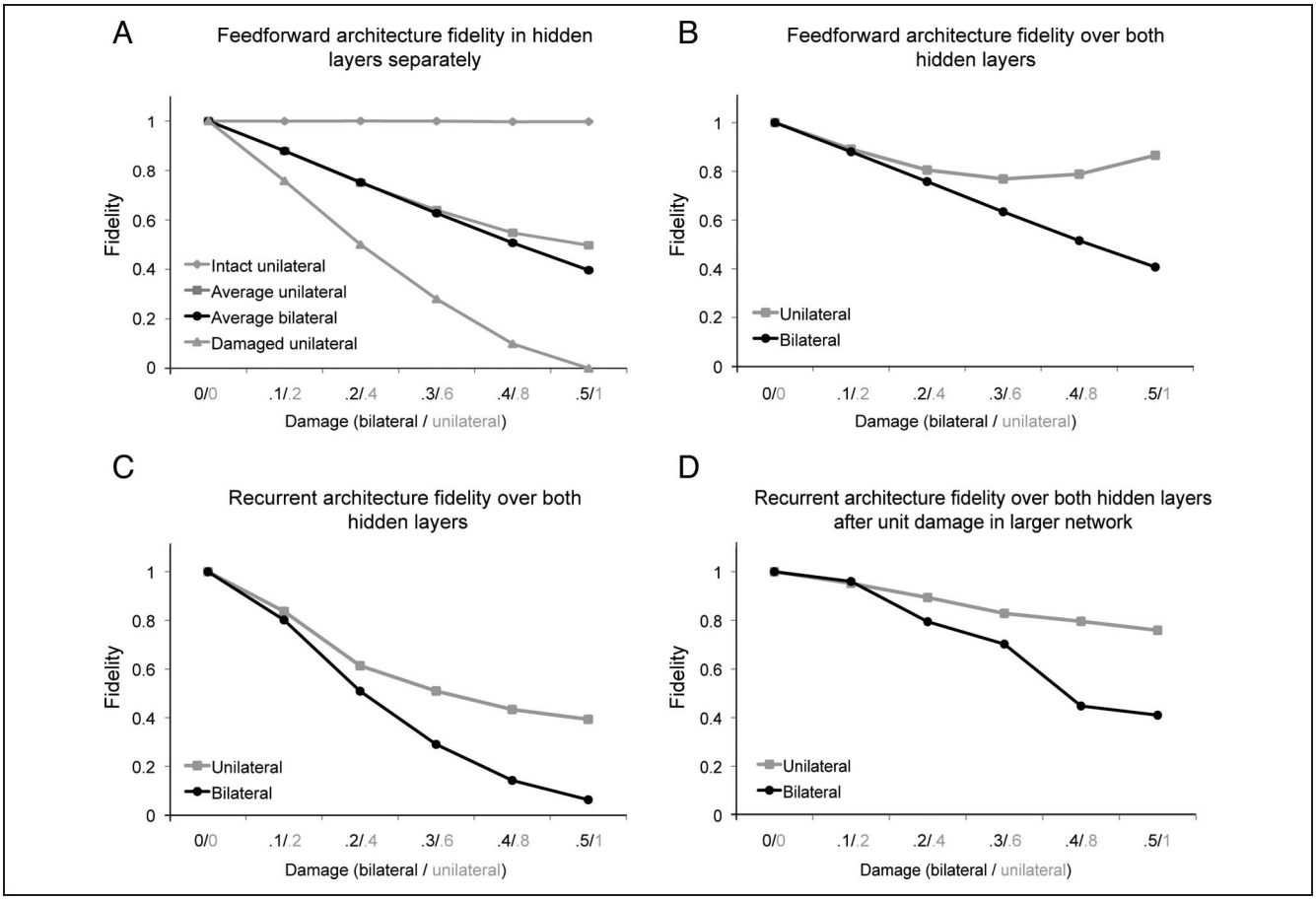


Figure 7. Fidelity of semantic representations as a function of damage. Fidelity was defined as the difference between mean diagonal and mean off-diagonal cosine values. The value of this difference was normalized by the value for the intact network, which therefore had a fidelity value of 1. (A) Fidelity in the damaged layer alone (damaged unilateral), the remaining, undamaged layer (intact unilateral); the average of these two (average unilateral); and the average of the fidelity across two bilaterally damaged demi-hubs matched for damage to the unilateral cases (average bilateral). (B) Fidelity measure applied directly over both hidden layers in the cases of unilateral and bilateral damage in the feedforward network. (C) Same as B but in the recurrent architecture. (D) Same as C, but after unit instead of connection damage. This network had 400 instead of 64 hidden units, as discussed in the text.

work, its own fidelity decreases, but the magnitude of its contribution to differences among the representations of different items also decreases, reducing its impact on overall fidelity, and therefore overall performance.

Simplified Mathematical Analysis

To see this effect in mathematical terms, we considered the effect of lesions on the cosine of the vectors of net inputs (excluding bias) to units in the hidden layer in the simplified feedforward architecture. Although activations, and not net inputs, were used in the network testing and fidelity calculations reported thus far, the net input is the signal that produces the meaningful differences among the hidden unit patterns produced by different items. The activation of each unit is a monotonic function of its net input, subject to perturbation by the Gaussian noise term added to each unit's net input. As the net inputs tend toward 0, all hidden unit activation values tend toward uninformative, uniformly distributed random values based on the Gaussian noise, causing the fidelity measure to fall to 0. Fidelity calculated using net inputs instead of activations yielded results similar to those in Figure 7A and 7B[1]. Based on these considerations, the analysis of net inputs we present below captures how the signal conveyed by the name or visual pattern for an item is affected by various levels of damage to connection weights.

To begin our examination of the net input cosines, let V_{v} be the net input across hidden layer units resulting from presentation of a particular object's visual features and let V_{n} be the net input across hidden layer units resulting from presentation of that same object's name. Each of these vectors can be broken up into the pattern of activation for the left half of the hidden units, V_{vl} and V_{nl}, and that for the right half, V_{vr} and V_{nr}. The cosine between the two complete vectors can be represented

$$\cos(V_{\mathrm{v}}, V_{\mathrm{n}}) = \frac{\sum_i V_{\mathrm{vl}i} V_{\mathrm{nl}i} + \sum_i V_{\mathrm{vr}i} V_{\mathrm{nr}i}}{\sqrt{\left(\sum_i V_{\mathrm{vl}i}^2 + \sum_i V_{\mathrm{vr}i}^2\right)\left(\sum_i V_{\mathrm{nl}i}^2 + \sum_i V_{\mathrm{nr}i}^2\right)}}.$$

Note that each of the two sums in the numerator corresponds to the dot product of the corresponding vectors, and (from the definition of the cosine, in Methods) the dot product of two vectors is equal to the cosine between the vectors times the product of the vectors' magnitudes. Similarly, each summation in the denominator corresponds to the square of the magnitude of the corresponding vector. Thus, the above equation can be rewritten

$$\cos(V_v, V_n) = \frac{||V_{vl}||||V_{nl}||\cos(V_{vl}, V_{nl}) + ||V_{vr}||||V_{nr}||\cos(V_{vr}, V_{nr})}{\sqrt{\left(||V_{vl}||^2 + ||V_{vr}||^2\right)\left(||V_{nl}||^2 + ||V_{nr}||^2\right)}}.$$

We call this the "magnitude-weighted cosine equation." We can now consider how unilateral versus bilateral damage affects this cosine by understanding how it affects the contributing magnitudes and cosines.

First, we consider in general terms the effect of damage to the weights coming into a common layer from two sources, corresponding to the name and the visual input to a pool of hidden units, on the net input cosines and magnitudes. The cosine reflects the degree to which the two net input vectors are similar. We consider the ideal case in which (before the lesion) the cosines on the left and right are both equal to 1 (i.e., the visual and name net inputs are identical). We also assume the left and right vectors have the same number of elements (same number of units on the left and on the right). We further assume that a lesion affects incoming weights from the name and visual units with equal probability p, so that the expected fraction of remaining incoming name and visual weights is $R = 1 - p$. These assumptions hold for the lesions we have performed on our networks. We can now examine how a lesion degrades this similarity relationship. As shown in Figure 8A, the average value after a lesion sparing proportion R of the weights (denoted by $< >_R$) of the cosine between two vectors is simply equal to R times the original value of the cosine:

$$< \cos(V_v, V_n) >_R = R\cos(V_v, V_n).$$

This "cosine shrinkage equation" applies to the effect of a lesion to incoming weights on the cosine over either half of the hidden units or to a uniform bilateral lesion over the weights coming to all of the hidden units.[2]

To see how the cosine is affected by a unilateral or, more generally, an unequal lesion, we must consider the effect of damage to weights on the magnitudes of the contributing net input vectors, as well as the effect on the contributing vectors' cosines. The average effect of a lesion leaving R remaining weights on the magnitude of a net input vector is $\sqrt{R}||V||$, where $||V||$ represents the vector's magnitude before the lesion.[3] For a lesion leaving the same proportion R of the weights contributing to V_v and V_n, we see that the average value of the product of the two vectors' magnitudes after the lesion $< ||V_v||||V_n|| >_R$ is equal to $R||V_v||||V_n||$, as shown in Figure 8B.

Now, we can consider separately the effect of a unilateral lesion versus a bilateral lesion. The remaining fraction of weights on each side, R_r and R_l, determine the relative weight of the right-side and left-side cosine in the value of the overall cosine (again, on each side, we treat the lesion as affecting the incoming name and visual weights equally). We assume that, before the lesion, the name and visual vectors have the same magnitude, and that the magnitude of the left half of each vector is the same on average as the magnitude of the right half of each vector. Effectively this means that the magnitudes of the four vectors $||V_{vl}||$, $||V_{vr}||$, $||V_{nl}||$, and $||V_{nr}||$ all have the same value M. Combining these assumptions with the finding above that $< ||V|| >_R = \sqrt{R}||V||$, replacing R with R_l and R with R_r for the proportions of connections remaining on the left and right sides, respectively, substituting into the magnitude-weighted cosine equation and simplifying, we find that

$$< \cos(V_v, V_n) >_{R_l,R_r} = \frac{R_l}{R_l + R_r} < \cos(V_{vl}, V_{nl}) >_{R_l} + \frac{R_r}{R_l + R_r} < \cos(V_{vr}, V_{nr}) >_{R_r}.$$

That is, after asymmetric damage, the relative weight of the left and right sides, w_l and w_r, are $w_l = R_l/(R_r + R_l)$ and $w_r = R_r/(R_r + R_l)$, respectively.

Table 2. Analysis of Variance for Figure 7 Results

	df	*F*	*p*
Feedforward Architecture Fidelity over Both Hidden Layers			
Damage	5, 45	1370.5	<.0001
Bilateral vs. mean unilateral	1, 9	683.52	<.0001
Damage * unilateral vs. bilateral	5, 45	411.84	<.0001
Recurrent Architecture Fidelity over Both Hidden Layers			
Damage	5, 45	861.61	<.0001
Bilateral vs. mean unilateral	1, 9	135.36	<.0001
Damage * unilateral vs. bilateral	5, 45	64.68	<.0001
Recurrent Architecture Fidelity over Both Hidden Layers after Unit Damage in Larger Network			
Damage	5, 45	71.09	<.0001
Bilateral vs. mean unilateral	1, 9	87.20	<.0001
Damage * unilateral vs. bilateral	5, 45	25.57	<.0001

Two-factor repeated-measures ANOVAs treat network initializations as the random effects factor, assessing overall effect of damage extent, effect of unilateral compared with bilateral damage, and the interaction between damage extent and the difference between unilateral and bilateral damage. Damage levels include zero to full unilateral damage over six increments (as in Figure 7).

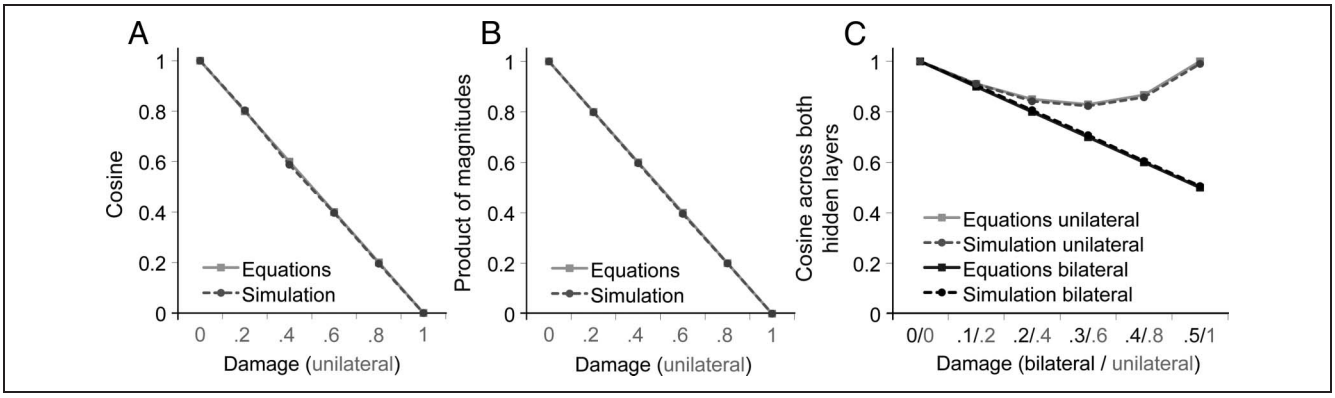


Figure 8. Effect of unilateral and bilateral connection weight removal on the cosine of the net input vectors at the hidden layer in the feedforward network. Analytic approximation (labeled "Equations"; solid lines) and actual values obtained in simulations (dashed lines). The cosine is calculated between the net input vector produced by inputting the name of an item with the cosine of the vector produced by the visual input pattern for the same item. (A) Effect of a lesion to weights coming into hidden units on one side on the cosine across the units on that side. (B) Effect of a lesion to weights on one side on the product of the magnitudes of the two vectors on that side. (C) Effect of a unilateral or bilateral lesion on the cosine of the net input vectors across both hidden layers. The curves labeled "equations" are based on the formulas given in the text. The simulation results were obtained by applying the following procedure to each of the 10 feedforward networks and averaging the results obtained for each network: At each lesion level, connections were removed from one or both hidden layers. Then all 48 names were presented, and the resulting hidden unit activation pattern was recorded for each. The same was done for all 48 visual patterns. There was no noise added to the net inputs in these simulations, and the bias term was not included in the net input value. The cosines (A and C) between each corresponding name and visual pattern as well as the product of the magnitudes of the members of each corresponding pair (B) were calculated. For each network, the 48 resulting cosines and the 48 resulting magnitudes were averaged.

In this equation, consider the effect of removal of some proportion of the connections coming into the right half of the hidden units. This will reduce the magnitudes of the net inputs to the right hidden units, and correspondingly it will therefore reduce the relative weight of the right-side cosine—in the extreme case of a total right-sided lesion, sparing all of the weights on the left ($R_l = 1, R_r = 0$), the right-side contribution goes to 0 and we are left only with the intact left side contribution, unaffected by the lesion. Crucially, in a bilaterally damaged network, where after the lesion V_{vl} and V_{nl} would have the same magnitude as V_{vr} and V_{nr}, the magnitude normalization would not result in domination by either side, and the distortion to the overall cosine would be the same as the equivalent distortion to each of the two cosines. These effects are depicted in Figure 8C, where we plot the overall cosine resulting from a pure one-sided lesion and the overall cosine resulting from an equal bilateral lesion, under the further simplifying assumption that the right- and left-sided cosines are both equal to 1 before the lesion. In the latter case, the cosine falls linearly with damage; in the former case, it actually follows a U-shaped pattern, decreasing then increasing again as damage weakens the relative contribution of units on the right hand side of the network. Once again the equation closely matches the result obtained by simulation.

The results of the analysis just described are closely approximated by the calculated values of the cosines after simulations of damage to the network, as shown by the agreement of the theoretical and simulation-based curves shown in Figure 8. (For the simulation results, the original values of the separate left, right, and overall cosines are not exactly equal to 1, because the net inputs from the name and visual units are not exactly identical. The results shown are normalized by the prelesion value of the relevant cosine term—left, right, or overall.) It is important to note that the expected (average) value of the cosine is not an exact predictor of accuracy in word-to-picture matching performance; the differences between cosines for targets and foils and the variability in these differences ultimately determines whether the correct or incorrect alternative is chosen. Nevertheless, the analysis provides a clear demonstration of the reason why a unilateral lesion has a relatively benign effect: Although the lesioned side may be grossly affected, its contribution to performance is reduced as a consequence of the lesion.

Fidelity and the Unilateral versus Bilateral Advantage in the Recurrent Network

Returning now to the more realistic recurrent architecture (from Effect of Lesions on Similarity Structure section), we can apply the fidelity analysis as before. Figure 7C shows the decreasing fidelity for unilateral and bilaterally damaged networks at increasing levels of damage. These curves resemble those for performance on word-to-picture matching and naming (Figure 4), indicating that these analyses provide a good account of the basis for the difference between unilateral and bilateral performance in those tasks.

It is interesting to consider why the fidelity falls more rapidly overall with damage in the recurrent network than in the feedforward network. This can be understood by noticing that, in the recurrent network, a lesion to

connections coming into hidden units on one side has a compounding effect. The first update to the activations of the hidden units in both demi-hubs is affected only by the damage to the connections from the pool that has inputs clamped, just as in the feedforward network. The distortion to unit activations is propagated via these recurrent connection weights, which themselves have been damaged, thereby increasing the distortion. Note that the propagated distortion affects both hubs to a degree, but because the extent of connectivity is greater within a hub than between hubs, the damaged hub has fairly little effect on the hub that was not damaged, contributing to the unilateral advantage.

Effect of Damage to Units in a Recurrent Network

Our simulations thus far have considered the effects of connection damage, but it might be argued that damage to whole units would be a more realistic model of damage in the brain, since degenerative illness, stroke, or resection might affect whole neurons or groups of neurons. To extend our findings to address this possibility, we ran damage simulations removing units instead of connections, where removal of a unit effectively removed all the connections into and out of that unit at once. We tested 10 networks using the recurrent network architecture, using a larger total number of hidden units (400 instead of 64) and longer training time (400,000 epochs). For simplicity, this simulation employed full connectivity within the left and the right side demi-hubs, no connections between the left and right sides, and no asymmetry in the output from hidden to verbal descriptors. We chose the larger number of hidden units because information is often not distributed very evenly across units in smaller networks and, consequently, removal of a unit can cause large network-specific changes that would not be expected in networks—like the brain—with much larger numbers of units.[4] The fidelity measure showed that unilateral damage was again less detrimental than bilateral damage and that the difference between the unilateral and bilateral damage increased with increasing amounts of unit damage (Figure 7D). These effects on fidelity carry over, as in other cases, to word-to-picture matching and naming.

The difference between unilateral and bilateral damage here arises entirely from the difference in recurrent connections within a hub versus between hubs. Recall that the net input to a unit is equal to a sum of terms, each of which is the activation of a sending unit multiplied by its connection weight to the receiving unit. Removal of units within a hub removes some of the terms from the net input to other units within the same hub, just as removal of connection weights among the units within a hub would. Because there is greater connectivity among hidden units within a hub than between hubs, the effect of unit removal is largely confined to units in the same hub. There can be indirect effects via recurrent influences propagating throughout the network, but as before, in the limit of removing all of the units on one side of the network, neither signal nor noise are propagated by this side of the network. In summary, a lesioned unit causes relatively focal severing of connections to other units in the same hub. As with connection damage, when lesioning is restricted to one side, that side's contribution is both weakened and distorted, allowing the intact side to dominate and thereby producing less of an overall deficit than a comparable amount of damage spread evenly across both sides.

Within versus Between Hemisphere Redundancy

We now consider a related but distinct possible explanation for the different consequences of unilateral and bilateral damage in humans and other animals. According to this explanation, the information contained in two homotopic areas is more redundant between than within each of the two areas, so if one is damaged, the information lost may still be available on the other side, but if both are damaged, some information lost from one side will also be lost from the other. Although we cannot rule out the possibility that such an account applies to lesions in the brain, it does not seem to be the cause of the differential effect of unilateral versus bilateral damage in our networks.

In our networks, there does not appear to be a force that produces greater redundancy of representation between versus within a hemisphere. Such a force certainly does not exist in our feedforward network (Figure 6). Because there are no recurrent connections and all input units are connected to all hidden units and all hidden units are connected to all output units, the division of the hidden layer into two pools has no effect whatsoever during training—the architecture is indistinguishable from that of a network with a single hidden layer. Thus, two units that might represent a particular feature are equally likely to be located on the same side of the network as on two different sides of the network. A consequence of this is that removing x units from each side has the same average effect on the fidelity of the representations in the network as removing $2x$ units from just one side.

To demonstrate that this is indeed the case for the feedforward network, we presented each item's name and visual input pattern, then computed the fidelity measure after removing $2x$ units from one side or x units from both sides. We found no reliable difference across 10 networks of unilateral versus bilateral removal across several values of the parameter x ($x = 0.1$, 0.3, and 0.5; smallest $p = .689$). We also carried out a similar analysis for two versions of the recurrent network: the network from Effect of Lesions on Similarity Structure section and the network with 400 hidden units. In both cases, we examined the fidelity of combined patterns

across the two hidden layers after the network was tested without damage, simply excluding units either unilaterally ($2x$ units on one side) or bilaterally (x units on both sides) in the analysis. As with the feedforward networks, for the same three values of x, we found no reliable effect of unilateral versus bilateral exclusion of units in either case (smaller network: smallest $p = .512$; larger network: smallest $p = .558$). These simulations are consistent with the view that there is no difference in redundancy of unit representation within versus between hemispheres in any of our networks.

Thus, although the greater redundancy between than within hemispheres account may seem in some ways similar to the account offered by our models, the accounts are in fact different. In the models, the removal of connection weights on just one side distorts the overall internal representation less than the removal of connection weights on both sides because the lesion to both sides distorts the representation on both sides, whereas the lesion on one side distorts the representation on only one side. Although signal distortion is an important part of the story, by itself it is incomplete. If damage results in signal distortion and downstream brain areas collect information from both hemispheres, the intact hemisphere would not necessarily override the noise arising from the damaged hemisphere. The insight from the current modeling work is that unilateral damage distorts the signal on that side but also results in a reduction in signal magnitude on that side. This magnitude reduction is what allows the intact side to dominate performance.

DISCUSSION

Across different species and multiple processing domains, unilateral damage can produce minimal effects whereas profound impairment is only observed after bilateral damage. Despite the fact that evidence for this difference in humans can be traced back for over half a century (to the famous case of patient HM) and longer in nonhuman primates (Klüver & Bucy, 1939; Brown & Schafer, 1888), there has been no computational investigation of the factors that underlie this difference until now (although there has been computational work on hemispheric specialization; e.g., Weems & Reggia, 2004; Lambon Ralph et al., 2001). We implemented a neural network model that simulates this difference and applied it to the domain of semantic memory, a domain in which there is a large body of relevant neuropsychological data. In line with the more general literature noted above, the patient data indicated that unilateral damage to the anterior temporal region generates only mild semantic impairment, whereas substantial semantic deficits follow bilateral ATL damage (as observed in SD and other bilateral diseases). Although matching for volume is difficult, the data suggest that unilateral damage is somehow inherently less deleterious than bilateral.

Our computational model exhibited just such a unilateral advantage. For symmetrically represented functions such as receptive comprehension (as measured by word-to-picture matching), unilateral damage only produced mild impairments, whereas the same quantity of simulated bilateral damage generated greater deficits at all levels of damage. For asymmetrically represented tasks such as picture naming, the simulations exhibited the expected laterality effect (lesions to the left demi-hub generated greater naming impairment than equivalent right demi-hub lesions; see also Lambon Ralph et al., 2001) while still showing an advantage, even for a unilateral lesion affecting the somewhat more dominant hemisphere, compared with equivalent damage spread across both hemispheres.

Part of the reason why performance after bilateral damage can be very poor is that it can affect all the units and connections whose function together determines network performance. As a result, as shown in Figure 4, the bilateral damage performance curves continue past the range possible for unilateral damage to yield very low levels of performance. It is probable that some of the patients with bilateral damage (e.g., those with severe SD) have similarly high amounts of atrophy.

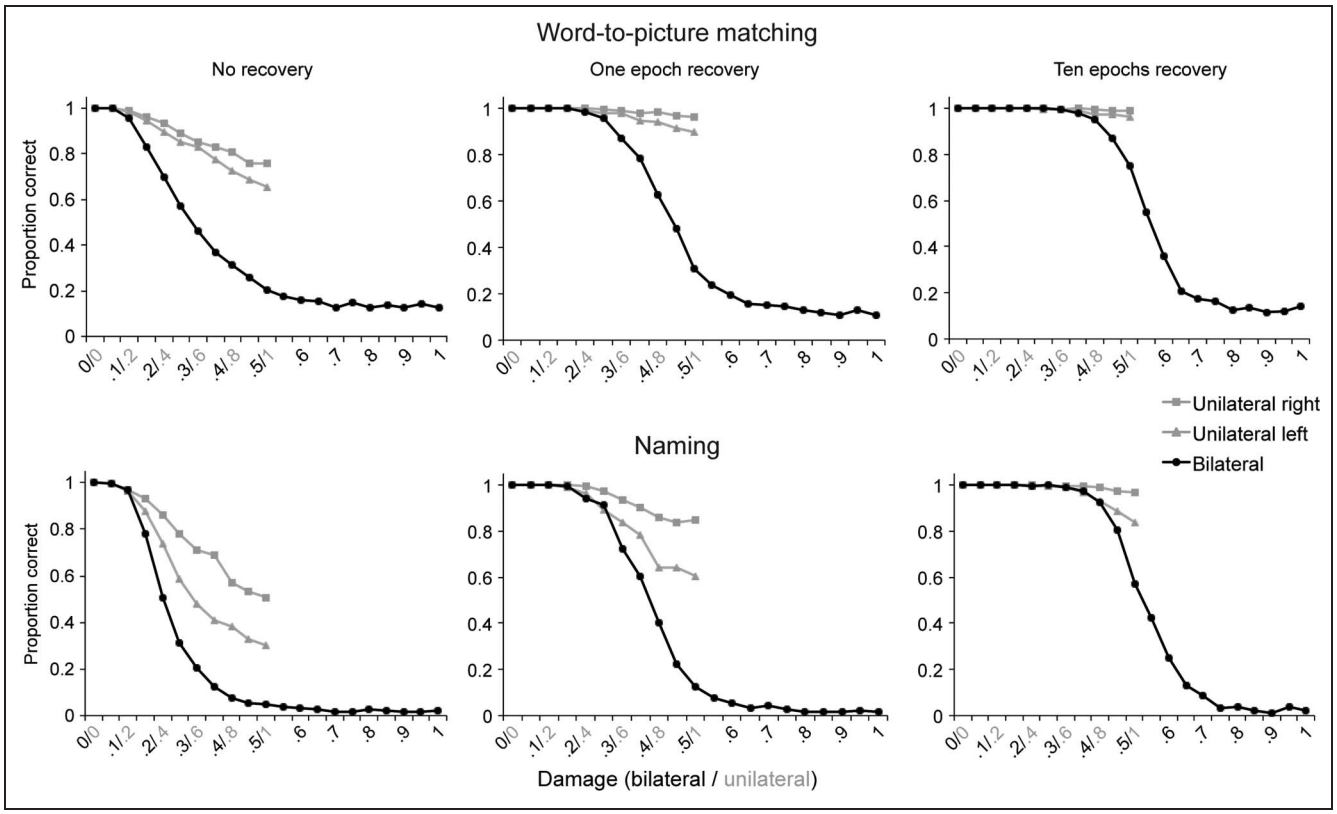
Our work indicates that there are also other important parts of the story. Within ranges where the model received equal amounts of damage unilaterally and bilaterally, performance after bilateral damage was much worse than after unilateral damage. The restriction of damage to one side of the model provides a considerable advantage over the same amount of damage distributed across both sides. When damage is restricted to one side, the other side can continue to perform much as it did when the network was fully intact, with relatively little disruption from the damaged side. Although it may be profoundly impaired, the damaged side contributes far less than the intact side because of the reduction in the variation in its activation from item to item, thereby allowing the intact side to dominate performance. When damage is applied to both halves of the network, processing is disrupted throughout, leading therefore to significantly greater behavioral impairment.

A mathematical analysis of the effects of damage demonstrated how distortion and signal magnitude interact in each hemisphere to produce different amounts of overall disruption in the unilateral and bilateral damage cases. This analysis combined with an understanding of the distributed representations and connectivity patterns of the network provides a clear picture of the causes underlying the differential effects of focal and diffuse damage.

Our results apply very generally to information processing in a highly interconnected system that is distributed bilaterally across the hemispheres of the brain. They may also be relevant to understanding effects of focal versus diffuse damage within a hemisphere in cases where a function is unilaterally represented over a brain region

of moderate extent. If in that case local connectivity among participating neurons is greater than long-range connectivity, then focal damage within the larger area would be expected to behave similarly to unilateral damage within our networks. More generally, although our model simulated the effects of unilateral versus bilateral ATL damage on semantic memory, our observations about the differential consequences of focal and diffuse damage are likely to be relevant to similar effects observed in many other domains.

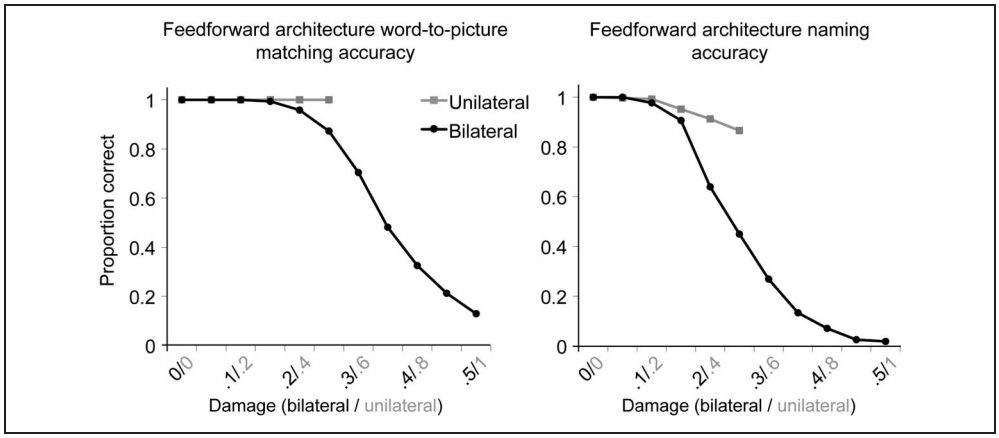
APPENDIX 1



Effects of unilateral and bilateral damage applied all at once (as opposed to progressively). Deterioration of performance is shown with no recovery (equivalent to the Figure 4 results), one epoch of recovery after the specified amount of damage, and 10 epochs of recovery after the specified amount of damage.

APPENDIX 2

Effects of unilateral and bilateral damage on naming and word-to-picture matching in the feedforward architecture. Performance is shown for the intact network and at 10 levels of damage, labeled by the proportion of connections removed from units in both hidden layers or one hidden layer. Deterioration of performance is shown with no recovery allowed.



Acknowledgments

We thank Ken Norman for helpful discussion and ideas about the simulation analyses. This work was supported by the National Science Foundation Graduate Research Fellowship under Grant DGE-0646086 to A. C. S. and by Medical Research Council program grants to M. A. L. R. (MR/J004146/1).

Reprint requests should be sent to Anna C. Schapiro, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08540, or via e-mail: schapiro@princeton.edu; Prof. Matthew A. Lambon Ralph, Neuroscience and Aphasia Research Unit, School of Psychological Sciences, Zochonis Building, Brunswick Street, Manchester, M13 9PL, UK, or via e-mail: matt.lambon-ralph@manchester.ac.uk or James L. McClelland, 344 Jordan Hall, Bldg 420, 450 Serra Mall, Stanford, CA, 94305, or via e-mail: mcclelland@stanford.edu.

Notes

1. In fact, using net inputs, we found that fidelity calculated for each half of the network separately almost exactly follows the fraction of connections remaining. For example, fidelity on either side after a 50% lesion is almost exactly .5. The average of left and right sided fidelity measures is thus the same for a bilateral lesion as it is for a unilateral lesion; there is no divergence at high damage levels as seen in Figure 7A. This is because net-input-based cosines decrease linearly with increasing damage while fidelity based on differences between activation-based cosines decrease non-linearly. The underlying cosines themselves actually increase at high levels of damage due to non-zero activation values.
2. Note that the cosine shrinkage equation is an empirical characterization of the average effect of a lesion leaving a remaining fraction R of the existing connection weights. When the incoming weights to a given hidden unit from active input units are correlated, the effect of a lesion on the cosine will be more benign, whereas if there is just one active input unit, or if the weights coming into each hidden unit from the ensemble of active input units are uncorrelated, the equation as given here should apply. It appears that the learning process results in connection weight values that are only weakly correlated in our simulations: First, as the figure shows, the equation closely matches the values obtained through simulations. Second, we estimated the correlation coefficient among the weights contributing to the net input to each hidden unit in each visual input pattern in each of the 10 feedforward networks (64 times 48 times 10 separate estimates) using the formula given in the Wikipedia article on Variance, section on Sum of correlated variables. The mean value of these estimates was 0.0334. While the value is significantly different from 0 ($t(9) = 20.06$, $p < .001$), it is clearly very small.
3. As with the effect of a lesion on the cosine, this is an empirical observation. In this case, as the incoming weights to each receiving unit from each active input unit became more correlated, the average effect of a lesion leaving R remaining weights on the length of a net input vector approaches $R||V||$, and not $\sqrt{R}||V||$. The observed value ($\sqrt{R}||V||$) is once again what is expected if there is just one active input unit, or if the weights coming into each hidden unit from the ensemble of active input units are uncorrelated.
4. Because there are far more connections than units, a lesion to connections tends to produce relatively more uniform effects, making smaller networks (which train far more quickly) sufficient for the connection damage simulations.

REFERENCES

Adolphs, R., & Tranel, D. (2004). Impaired judgments of sadness but not happiness following bilateral amygdala damage. *Journal of Cognitive Neuroscience, 16,* 453–462.

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: Evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex, 20,* 2728–2738.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38,* 1207–1215.

Braitenberg, V., & Schüz, A. (1998). *Cortex: Statistics and geometry of neuronal connectivity*. New York: Springer.

Brown, S., & Schafer, E. (1888). An investigation into the functions of the occipital and temporal lobes of the monkey's brain. *Philosophical Transactions of the Royal Society of London, 179,* 303–327.

D'Esposito, M., Cooney, J. W., Gazzaley, A., Gibbs, S. E., & Postle, B. R. (2006). Is the prefrontal cortex necessary for delay task performance? Evidence from lesion and fMRI data. *Journal of the International Neuropsychological Society, 12,* 248–260.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., et al. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage, 11,* 589–600.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120,* 339–357.

Gee, D. G., Biswal, B. B., Kelly, C., Stark, D. E., Margulies, D. S., Shehzad, Z., et al. (2011). Low frequency fluctuations reveal integrated and segregated processing among the cerebral hemispheres. *Neuroimage, 54,* 517–527.

Giovagnoli, A. R., Erbetta, A., Villani, F., & Avanzini, G. (2005). Semantic memory in partial epilepsy: Verbal and non-verbal deficits and neuroanatomical relationships. *Neuropsychologia, 43,* 1482–1492.

Glosser, G., Salvucci, A. E., & Chiaravalloti, N. D. (2003). Naming and recognizing famous faces in temporal lobe epilepsy. *Neurology, 61,* 81–86.

Heffner, H. E., & Heffner, R. S. (1986). Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *Journal of Neurophysiology, 56,* 683–701.

Hermann, B. P., Seidenberg, M., Haltiner, A., & Wyler, A. R. (1995). Relationship of age at onset, chronologic age, and adequacy of preoperative performance to verbal memory change after anterior temporal lobectomy. *Epilepsia, 36,* 137–145.

Hermann, B. P., Wyler, A. R., Somes, G., Dohan, F. C., Jr., Berry, A. D., III, & Clement, L. (1994). Declarative memory following anterior temporal lobectomy in humans. *Behavioral Neuroscience, 108,* 3–10.

Hinton, G. E., & Sejnowski, T. J. (1983). *Learning and relearning in Boltzmann machines*. Cambridge, MA: MIT Press.

Keidel, J. L., Welbourne, S. R., & Lambon Ralph, M. A. (2010). Solving the paradox of the equipotential and modular brain: A neurocomputational model of stroke vs. slow-growing glioma. *Neuropsychologia, 48,* 1716–1724.

Klüver, H., & Bucy, P. C. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. *Archives of Neurology and Psychiatry, 42,* 979–1000.

Lambon Ralph, M. A., Cipolotti, L., Manes, F., & Patterson, K. (2010). Taking both sides: Do unilateral, anterior temporal-lobe lesions disrupt semantic memory? *Brain, 133,* 3243–3255.

Lambon Ralph, M. A., Ehsan, S., Baker, G. A., & Rogers, T. T. (2012). Semantic memory is impaired in patients with

unilateral anterior temporal lobe resection for temporal lobe epilepsy. *Brain, 135,* 242–258.
Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: Neuropsychological abstract evidence and a computational model. *Journal of Cognitive Neuroscience, 13,* 341–356.
Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex, 19,* 832–838.
Lewis, J. D., Theilmann, R. J., Sereno, M. I., & Townsend, J. (2009). The relation between connection length and degree of connectivity in young adults: A DTI analysis. *Cerebral Cortex, 19,* 554–562.
Li, H., Matsumoto, K., & Watanabe, H. (1999). Different effects of unilateral and bilateral hippocampal lesions in rats on the performance of radial maze and odor-paired associate tasks. *Brain Research Bulletin, 48,* 113–119.
Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron, 38,* 487–497.
Martin, R. C., Sawrie, S. M., Roth, D. L., Gilliam, F. G., Faught, E., Morawetz, R. B., et al. (1998). Individual memory change after anterior temporal lobectomy: A base rate analysis using regression-based outcome methodology. *Epilepsia, 39,* 1075–1082.
Milner, B., & Penfield, W. (1955). The effect of hippocampal lesions on recent memory. *Transactions of the American Neurological Association*, *80*, 42–48.
Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., et al. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain, 133,* 3256–3268.
Nestor, P. J., Fryer, T. D., & Hodges, J. R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage, 30,* 1010–1020.
Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., et al. (2007). Temporal lobe lesions and semantic impairment: A comparison of herpes simplex virus encephalitis and semantic dementia. *Brain, 130,* 1138–1147.
Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30,* 1025–1040.
Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8,* 976–987.
Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 20137–20141.
Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia, 48,* 1336–1342.
Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111,* 205–235.
Rohrer, J. D., Warren, J. D., Modat, M., Ridgway, G. R., Douiri, A., Rossor, M. N., et al. (2009). Patterns of cortical thinning in the language variants of frontotemporal lobar degeneration. *Neurology, 72,* 1562–1569.
Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, Psychiatry, 20,* 11–21.
Seidenberg, M., Hermann, B., Wyler, A. R., Davies, K., Dohan, F. C., Jr., & Leveroni, C. (1998). Neuropsychological outcome following anterior temporal lobectomy in patients with and without the syndrome of mesial temporal lobe epilepsy. *Neuropsychology, 12,* 303–316.
Sharp, D. J., Scott, S. K., & Wise, R. J. (2004). Retrieving meaning after temporal lobe infarction: The role of the basal language area. *Annals of Neurology, 56,* 836–846.
Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioral Neurobiology, 2,* 167–182.
Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature, 383,* 254–256.
Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon Ralph, M. A. (2010). The inferior, anterior temporal lobes and semantic memory clarified: Novel evidence from distortion-corrected fMRI. *Neuropsychologia, 48,* 1689–1696.
Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience, 22,* 1083–1094.
Visser, M., & Lambon Ralph, M. A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of Cognitive Neuroscience, 23,* 3121–3131.
Warden, C. J., Barrera, S. E., & Galt, W. (1942). The effect of unilateral and bilateral frontal lobe extirpation on the behavior of monkeys. *Journal of Comparative Psychology, 34,* 139–171.
Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27,* 635–657.
Weems, S. A., & Reggia, J. A. (2004). Hemispheric specialization and independence for word recognition: A comparison of three computational models. *Brain and Language, 89,* 554–568.
Welbourne, S. R., & Lambon Ralph, M. A. (2005). Exploring the impact of plasticity-related recovery after brain damage in a connectionist model of single-word reading. *Cognitive Affective & Behavioral Neuroscience, 5,* 77–92.
Wilkins, A., & Moscovitch, M. (1978). Selective impairment of semantic memory after temporal lobectomy. *Neuropsychologia, 16,* 73–79.