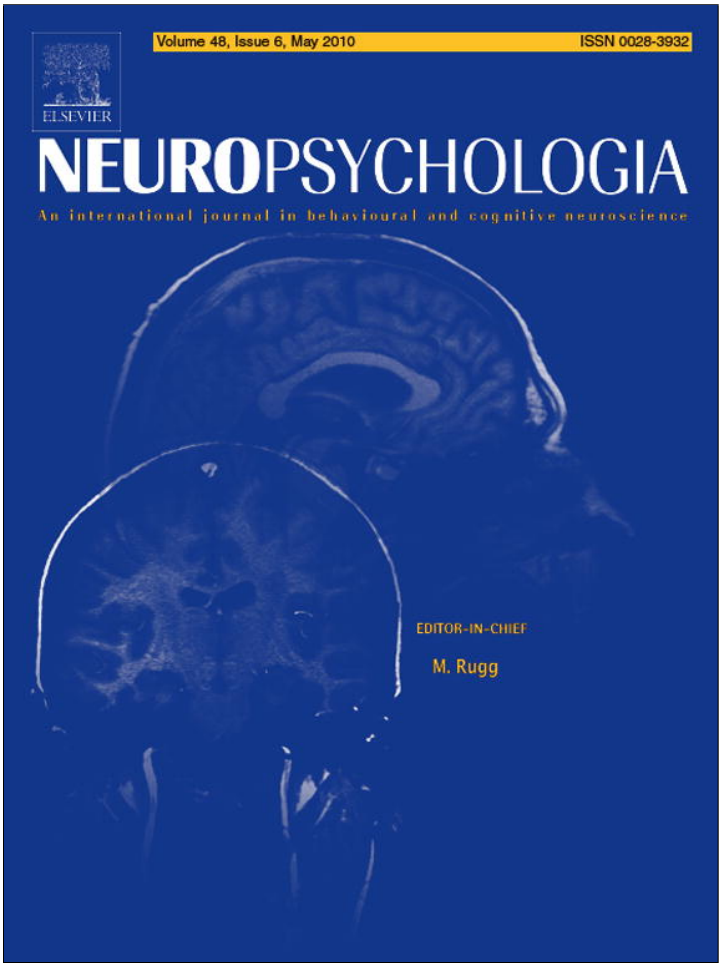
Volume 48, Issue 6, May 2010
ISSN 0028-3932
ELSEVIER
NEUROPSYCHOLOGIA
An international journal in behavioural and cognitive neuroscience
EDITOR-IN-CHIEF
M. Rugg

Neuropsychologia 48 (2010) 1716–1724

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Solving the paradox of the equipotential and modular brain: A neurocomputational model of stroke vs. slow-growing glioma

James L. Keidel, Stephen R. Welbourne, Matthew A. Lambon Ralph*

Neuroscience and Aphasia Research Unit (NARU), School of Psychological Sciences (Zochonis Building), University of Manchester, Brunswick Street, Manchester M13 9PL, UK

ARTICLE INFO

Article history:
Received 21 August 2009
Received in revised form
31 December 2009
Accepted 15 February 2010
Available online 24 February 2010

Keywords:
Language
Plasticity
Entrenchment
Glioma
Stroke
Computational modelling

ABSTRACT

In acute brain damage (e.g., stroke), patients can be left with specific deficits while other domains are unaffected, consistent with the classical 'modular' view of cortical organization. On this view, relearning of impaired function is limited because the remaining brain regions, tuned to other domains, have minimal capacity to assimilate an alternative activity. A clear paradox arises in low-grade glioma where an even greater amount of cortex may be affected and resected without impairment. Using a neurocomputational model we account for the modular nature of normal function as well as the contrasting types of brain insult through the interaction of three computational principles: patterns of connectivity; experience-dependent plasticity; and the time course of damage. This work provides support for a neo-Lashleyan view of cortical organization.

1. Introduction

A critical aim of behavioural neurology, neuropsychology and now cognitive neuroscience is to understand how various behaviours are based on the function of different brain regions. Extending back to the work of the nineteenth century neurologists, scientists have attempted to relate impaired performance, sometimes reflecting specific behavioural dissociations, to the area of underlying brain damage. Often three simplifying assumptions are made in this process: (a) *modularity*—that complex behaviours are made up from cognitively separable processing steps which can be independently impaired by brain damage; (b) *transparency*—that a patient's impaired performance directly reflects the area of brain damage; and (c) *subtractivity*—following from the assumption of modularity, that the patient's impaired performance reflects the simple, independent subtraction of a specific function from the normal system (i.e., without resultant changes in brain anatomy or function). When reviewing the neurological and neuroscience literature, however, an apparent paradox arises; some types of neurological damage and neuroscience data align closely with this classical "modular" view while others point to a more complex relationship between brain and behaviour, in which a considerable degree of neuroplasticity plays a key role.

There is certainly no lack of evidence for relatively strict structure–function correspondences in the human brain. For example, averaging across individual subjects for group analysis of neuroimaging data implicitly relies on such correspondences (Price & Friston, 2002). Indeed, most human behaviours exhibit reproducible loci of activation in response to stimuli of a given class. Thus speech perception typically activates a bilateral network centred around primary auditory cortex, visual processing reliably engages a well-defined bilateral network of occipital and temporal areas, and motor tasks predictably activate cortex in keeping with the layout of the 'homunculus' represented in M1 (Rao et al., 1995). This tight relationship between function and structure is then reinforced when neurological and neuroimaging data are combined; brain damage or transcranial magnetic stimulation to these areas often leads to relatively well-circumscribed deficits/effects related to that area's function in the typical, undamaged case.

While the modular view has proved extremely valuable for neurological, neuropsychological and neuroscientific theory, the past three decades have witnessed demonstrations of lifelong neural plasticity far beyond what might have been imagined based on the classical view alone. Animal studies have exhibited a startling degree of equipotentiality in the early stages of development as well as the continuing ability in adult animals to restructure cortical maps in response to neural degradation (Bao, Chang, Davis, Gobeske, & Merzenich, 2003). In humans, altered cortical represen-

* Corresponding author. Tel.: +44 0161 275 2551/2581; fax: +44 0161 275 2873.
E-mail address: matt.lambon-ralph@manchester.ac.uk (M.A. Lambon Ralph).
URL: http://www.psych-sci.manchester.ac.uk/naru/ (M.A. Lambon Ralph).

0028-3932/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2010.02.019

tations have been demonstrated in response to training on a new motor task, pitch discrimination and identification of novel visual objects (Bosnyak, Eaton, & Roberts, 2004; Draganski et al., 2004; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999).

But if the brain is so plastic, why is the prognosis for stroke patients so poor? Very recently, a series of papers by Duffau and his colleagues on the behavioural sequelae of low-grade glioma (LGG) has provided new insights into this apparent paradox (Desmurget, Bonnetblanc, & Duffau, 2007; Duffau, 2005, 2006). These tumours gradually destroy large swathes of cortex (including documented cases in Broca's area, supplementary motor cortex, and the temporal lobe) over the course of years without inducing significant behavioural impairment. In most cases, within a few months of tumour resection the patient returns to a normal socioprofessional life. Compare this to stroke, in which comparatively smaller lesions often lead to irreversible deficits even after a long period of limited, partial recovery. While the different time course between the two types of lesions provides a clue to the root causes of these contrasting outcomes, a complete account requires an elucidation of the underlying neurocomputational mechanisms. We propose here three key factors that determine the success of relearning after brain insult:

(1) *The age at which damage occurs*. While the commonsense notion that 'earlier is better' does not always hold true, especially in the case of congenital deficits affecting neural development generally, it is nonetheless clear that outcomes are often better when damage is sustained earlier in development. Studies of children who have undergone hemispherectomy to control intractable epilepsy represent perhaps the most extreme example of this capability. While this surgery and the seizures that precede and sometimes follow it often result in significant cognitive impairment, a number of studies have demonstrated surprisingly high language function in these patients. Liegeois, Connelly, Baldeweg, and Vargha-Khadem (2008) reported on language performance in a cohort of 30 hemispherectomy patients and found that patients whose surgery involved the right hemisphere and whose damage occurred postnatally had verbal IQs in the low normal range. Another large-scale study assessed postsurgical spoken language outcomes in 43 hemispherectomy patients and found highly varying outcomes including many patients with complete mature grammars (Curtiss, de Bode, & Mathern, 2001). Muller et al. (1999) directly compared groups with early and late left hemisphere lesions and found PET evidence for increased cortical reorganization in those participants who suffered damage before the age of five. Such cases provide a stark contrast to the outcomes observed after significant damage to the brain in adults.

(2) *The time course of the damage*. The differences observed in LGG and stroke provide strong evidence for the central role of this factor in determining the end result of neural reorganization. In the case of LGG, the damage takes place continuously but slowly over years (the typical increase in diameter is 4 mm/year), allowing other neural regions to assume the role previously filled by the deteriorating cortex. In stroke, the core damage is nearly instantaneous, penumbra function is lost over a matter of days and the compensatory action of other cortical regions is often insufficient to support normal processing. We hypothesize that the ability of infiltrated cortex to guide the development of new cortical networks that assume responsibility for the deteriorating function allows for the surprisingly positive outcomes observed in LGG. In the case of stroke, this may not be possible, as the knowledge that was encoded in the damaged cortex rapidly becomes inaccessible to other cortical regions. In the case of slowly progressing damage, however, the decaying cortex can maintain representations close enough to those of the normal state to guide restructuring successfully.

(3) *The pattern of connectivity of the damaged cortex*. A third key principle is that the brain is not fully interconnected but that neighbouring neurons are connected with a high probability via intracortical connections, whereas distant neurons are connected more sparsely via white matter intercortical connections (as indicated by histological studies: Young, Scannell, & Burns, 1995). We hypothesize that this partial and regionally specific pattern of white matter connectivity, in concert with the pattern of connectivity exhibited by primary sensory cortices, plays a central role in the establishment of modularity in the adult human brain. That is, in typically developed individuals some basic regional specification necessarily follows from the fact that initial cortical processing of visual input relies on the thalamocortical afferents from the lateral geniculate nucleus to primary visual cortex (V1) in the occipital lobe, while auditory input first arrives at Heschl's gyrus in the temporal lobe. Unsurprisingly, then, auditory association areas are found in the superior temporal gyrus lateral to A1, visual association areas in occipital regions surrounding V1, and audiovisual processing (e.g., reading) in areas close to the junction of the occipital and temporal lobes.

To explore the neurocomputational basis of modular organization in the normal mature brain and the contrasting effects of different types of brain damage, we trained a series of parallel distributed processing (PDP) neural network models that embodied and tested these three neurocomputational principles. Many computational models simulate acquisition of a single behaviour (e.g., reading aloud, naming, etc.). In order to demonstrate modularity and subsequent behavioural dissociations after damage, however, the models used in this study were trained to perform two different quasi-regular mapping tasks. We then compared the differential effects of acute or gradually incremental damage. To test this notion, we implemented full intra-connectivity within two half 'subnetworks' (a computational analogy to different two different brain regions) but only partial inter-connectivity between them. Having confirmed this hypothesis in the trained model, we then damaged one subnetwork in two different ways. In the LGG simulation, the weights were slowly reduced to zero through addition of a decay term. In the stroke simulation, these same weights were destroyed instantaneously. Based on the principle that information encoded in the brain is continually updated through generalized learning/plasticity in response to continued life experiences, the damaged simulations were re-exposed to their learning environment to allow for experience-dependent plasticity-related recovery (Welbourne & Lambon Ralph, 2005, 2007). The time course and endpoints of recovery in both simulations closely reflected those observed in stroke and tumour patients.

2. Experiment 1

2.1. Method

The architecture used in these simulations consisted of two parallel three-layer feedforward networks, with 125 hidden units and 50 units in the input and output groups, as depicted in Fig. 1. Pilot simulations established that this number of hidden units within a single "subnetwork" was sufficient for learning both computational tasks (if trained from scratch on both tasks). As a consequence we knew for certain that the damaged models had sufficient computational resources to accommodate both activities. However, as we will go on to demonstrate, changes in the learning principle parameters (age of damage and speed of damage) altered the ability of this "fully resourced" model to recover function. Following the differential connectivity between neighbouring vs. long-distance neurons (see Section 1), groups of units within each subnetwork were fully connected while sparse cross-connections (each unit was connected to approximately 30% of the units in the downstream layer) linked into and out of the other subnetwork's hidden layer. Each subnetwork was

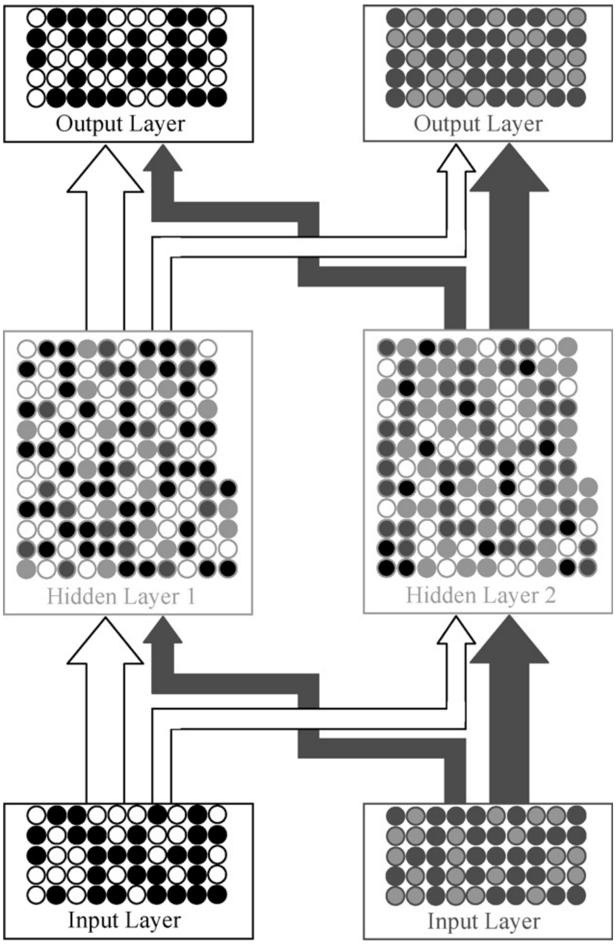

Fig. 1. The architecture of the neurocomputational model. The neurocomputational model was trained to perform two unrelated tasks simultaneously. Task 1 consisted of a series of binary patterns presented to input layer 1 and the model was trained to produce the corresponding pattern on output layer 1, via all intermediate (hidden) units. Likewise task 2 required the correct pattern to be produced on output layer 2 given the pattern on input layer 2, via the hidden units. No direct constraint was placed on the hidden units in terms of which task they could participate in. The only difference built into the model was the connectivity pattern. Input, hidden and output layers 1 were fully connected. The same was true for input, hidden and output layers 2. There were, in addition, sparser (30%) connections between input layer 1 to hidden layer 2 to output layer 1, and input layer 2 to hidden layer to output layer 2. All connections were initiated with small random values and then these were adjusted using the backpropagation learning algorithm so that the model was able to perform both tasks without error. Once trained, two different kinds of damage to hidden layer 1 were compared: acute (instantaneous) removal of the connections to and from the layer vs. a gradual reduction (decay) of the strength of these same connections to zero. Hidden layer 2 and its connectivity were not damaged. In order to allow for experience-dependent plasticity-related changes, the learning algorithm continued to be applied after both types of damage.

trained simultaneously on one of two independent tasks so that we could compare performance on each of them both at the end of initial training (to test for emergent modularity in the normal "adult" model) and at different points post damage (in the case of simulated stroke) or during damage (in the case of simulated LGG). The two tasks were designed to be independent – i.e., the patterns used in one task did not relate in any systematic way to the patterns used for the other task – thus removing the possibility of direct "cross-talk" of the two tasks that would then complicate the interpretation of the model in each of its phases. Specifically, the input sets for each task were comprised of separate groups of 100 random 50-bit input vectors. The targets for these inputs were created by randomly flipping approximately 30% of these bits; thus each input → target mapping consisted of a more predictable part (the cases where, say a 1 on input unit 39 corresponded to a 1 on output unit 39) and a less predictable part (a case where a 1 on input 39 corresponded to a 0 on output 39). Note that these targets remained fixed throughout training—bits were not flipped on a per-trial basis. On each trial one of the 100 input patterns for subnetwork 1 was presented to input layer 1 and at the same time one of the input patterns for subnetwork 2 was applied to input layer 2. All possible combinations of input patterns were presented, yielding $100^2 = 10{,}000$ training patterns. The model was trained to produce the correct output for each pattern set on the two output layers.

The simulations were generated and trained using the LENS neural network simulator (v2.63 from http://tedlab.mit.edu/~dr/Lens/) with the following parameters. All simulations used online learning (i.e., weight changes were made after each trial), logistic activation and backpropagation of cross-entropy error, with a learning rate of 0.01. To simulate age-related decrease in neural plasticity, models were trained with a linearly increasing 'entrenchment factor', instantiated as the inverse of a logistic cost function (the logit function) on both the output layers of the form:

$$\text{Cost} = \frac{(\ln(o_j) - \ln(1 - o_j)) \times C_s}{\ln(C_p)} \tag{1}$$

where o_j refers to the output of unit j, C_s (cost strength) is a constant that controls the magnitude of the cost function, and C_p (cost peak) is the output value that incurs the largest resistance (in this case, the middle of the logistic output function, 0.5). This value is then added to the weight change for the given trial. At the beginning of training the cost strength was set to 0, and increased to 4.0 at the end of training. The effect of this entrenchment factor is to favour weight changes that lead to binary outputs, independent of the target value. This continual increase in the cost strength provides a preference for maintenance of stored knowledge over significant change in the model's internal representations: see Section 6 and Appendix for further details. Performance on each trial was scored as the percentage of units whose activation was on the correct side of 0.5 for each target; this score was then averaged across the 10,000 patterns. After 200,000 trials, error had reached the minimum asymptote and training was concluded.

Simulation of stroke was performed by complete and simultaneous removal of one hidden layer—while this is a severe manipulation, we wished to observe model behaviour at the extremes of damage (the effects of partial damage are explored in Section 5, below). Following this intervention, the model was re-exposed to the training environment for 2 million trials, which allowed recovery to reach asymptote (mimicking the partial recovery shown by some stroke patients after damage). It should be noted that we included a simulated "recovery" phase that was much longer than the original developmental phase. We did this in order to provide the stroke simulation with an unfettered opportunity to recovery and to reach an asymptote (i.e., extend the recovery to an extreme position beyond the life expectancy of a human). This allowed us to test whether the two simulations would demonstrate a substantial difference not only during recovery but also at asymptote. In real life, the time post damage would represent a reduced duration on the recovery curve—at which point, as we will report below, there is an even greater difference between the two types of simulation than at asymptote.

In contrast to the acute and simultaneous damage of stroke, the tumour simulation involved the imposition of a high level of weight decay to all the links entering and exiting one hidden layer, with the learning rate of these links set to 0. The effect of this alternative form of damage is that the value of each weight is gradually reduced to zero rather than being forced to zero instantaneously (as in the stroke simulation). Thus, although the end point of each form of damage is the same (deletion of the weight values that code knowledge/information in this kind of neural network model), the critical difference is the time over which this end point is reached. It should be noted that these two forms of damage were applied to exactly the same model, with the same "adult"-level of entrenchment. Therefore, if any differences emerged from the two kinds of damage then these must reflect the form of damage rather than any other uncontrolled parameter.

3. Results

3.1. Emergent modularity in the adult brain

After training, the network (depicted in Fig. 1) performed the two separate tasks with 100% accuracy. In neurocomputational models of this type (which most commonly have full connectivity), the information (weight values) to complete multiple tasks would be spread across the entire network in a completely homogenous fashion. As noted in Section 1, our working hypothesis was that modular processing would follow from restricting connectivity. The functioning of units within such models is shaped by the tasks/inputs/outputs that they are connected to. With full connectivity, the units are equipotential. In contrast, restricted connectivity results in increasing specialization (Plaut, 2002). This hypothesis was tested by severing the sparser cross-connections connecting the two subnetworks. Despite the fact that this is equivalent to the removal of ~25% of the model's representational capacity (as measured by number of links), this intervention had no effect on the performance of either subnetwork (performance remained at 100% for both tasks). An analysis of the strength of acti-

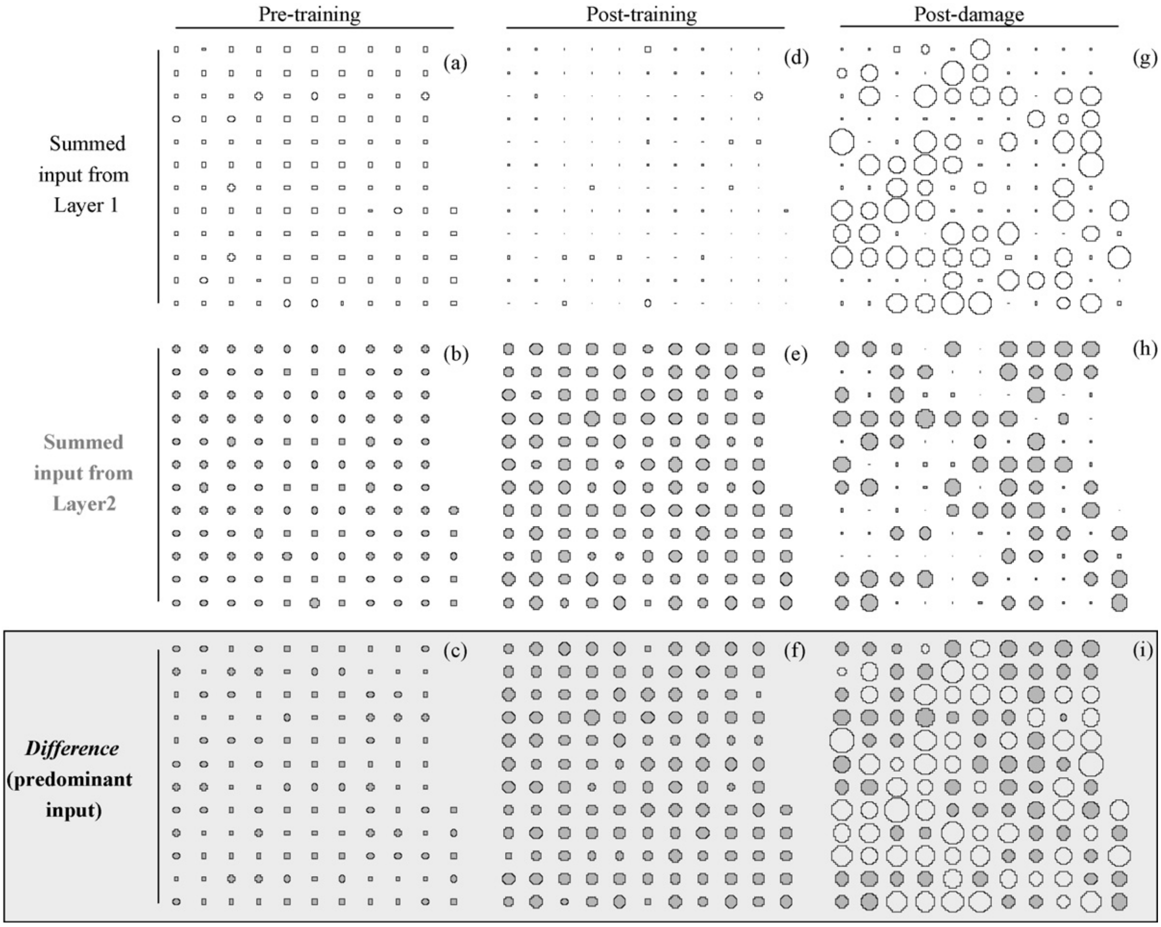

Fig. 2. Comparison of summed net input to the undamaged hidden layer 2 before training, after training (normal development) and after slow weight decay to hidden layer 1 (low-grade glioma). This series of nine bubblegrams depicts the strength of input into each unit within the undamaged hidden layer 2 at different time points. In the top row, the total input from the more sparsely connected input layer 1 is shown. The middle row shows the total input from the fully connected input layer 2. The bottom row shows the difference between the total input from each input layer and thus reflects the dominant input. Prior to any training (left column) the units in hidden layer 2 have small inputs from both input layers 1 (panel a) and 2 (panel b). Due to the relative sparsity of the connection from input layer 1, the magnitude of the input from input layer 1 is smaller and thus layer 2 is slightly more dominant (panel c). The emergence of the quasi-modular behaviour of the model after training is shown in the middle column. Specifically, the magnitude of input from layer 1 drops (compare panels a and d) to minimal values while the input from layer 2 grows (compare panels b and e). As a result, the functioning of hidden layer 2 is almost completely dominated by input layer 2 (panel f)—i.e., it becomes a quasi-modular system for task 2. The system is only *quasi*-modular, however, because under the right circumstances – in this case slow weight decay of the connected to hidden layer 1 – the function of hidden layer 2 can change (right column). Around half of the units remain dominated by input from input layer 2 (panel h) and, if anything, the magnitude of the input actually increases in order to maintain performance on task 2 with only half the number of contributing units. Quite strikingly, the other half (those with the greatest connectivity to input layer 1) completely change their functioning in favour of input layer 1 (panel g). When taken together (panel h) it becomes clear that both task 1 and task 2 are supported by the functioning of the units of hidden layer 2, with around half relatively dedicated to each task.

vation from each input layer to the hidden layer also confirmed the same modular functioning (see Fig. 2). Thus, following the pattern of connectivity, each subnetwork had become relatively specialized for one task—i.e., modular, despite the fact that the “latent” cross-connections allowed for the possibility of a mixed contribution to both tasks from each subnetwork. The next step was to test the response of this modularized network to acute (stroke) vs. slowly progressive (LGG) simulated damage.

3.2. Stroke simulation

Stroke was simulated through the complete and instantaneous deletion of the hidden layer in one subnetwork, rendering all links entering and exiting that layer non-functional. Immediately following this lesion, activations on the output layer of the damaged subnetwork were effectively random and unrelated to the input pattern—i.e., like many patients, immediately after their stroke, performance on the affected task was at floor level. After the experience-dependent recovery period, the stroke models' performance increased to 69.7% (Fig. 3), which mimics the partial recovery demonstrated by many patients (to variable degrees) in the first few months following their stroke (Wade, Hewer, David, & Enderby, 1986; Welbourne & Lambon Ralph, 2005, 2007). Importantly, as observed in stroke, the initial post-lesion recovery was followed by a chronic phase in which little or no additional relearning took place; the performance curve for stroke in Fig. 3 reaches asymptote at 70%, and remains there even when the model is run for an additional 10 million trials. At no point did performance on the unaffected task fall below 100% – that is to say, following expectations from a quasi-modular system, the model – like stroke patients – demonstrated a neuropsychological dissociation between the two tasks as well as partial recovery of the affected domain.

3.3. LGG simulation

In order to capture the slow, gradual infiltration of LGG (Mandonnet et al., 2003), weight decay was added to all links entering and exiting one hidden layer, causing a gradual reduction in the influence of this hidden layer on the output of the lesioned subnetwork. The result of this gradual damage was quite unlike

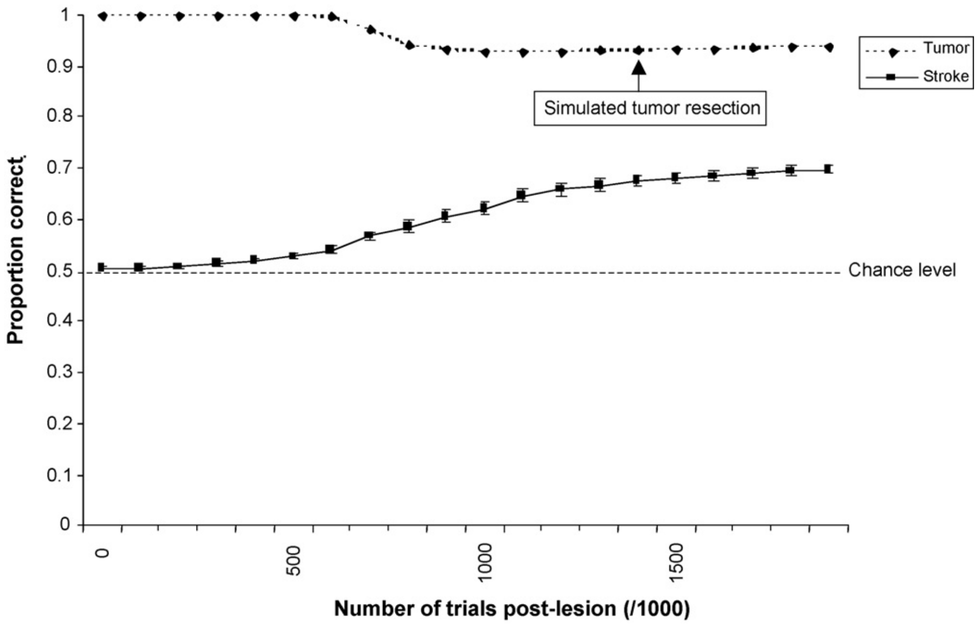

Fig. 3. Relative performance of the stroke and tumour simulations on task 1. Performance of the model on the affected task (task 1) is plotted against the number of post-lesion training trials. There was no effect on the undamaged task 2, so this is not plotted (see text). Accuracy is measured in terms of the proportion of units with the correct target value. Thus 50% correct corresponds to chance level performance. Simulations of acute stroke produced a catastrophic loss of function in the immediate phase followed by partial recovery. Simulated LGG only led to minimal reduction in accuracy on task 1. The arrow at 1.4 M trials denotes that, at this point in the tumour model, the entire hidden layer 1 could be removed with no effect on task 1.

the instantaneous stroke simulation: despite considerable, cumulative damage to this subnetwork, performance on the associated task only dropped by a minimal amount (accuracy never fell below 93%). This relative preservation of task performance (which had been decimated in the stroke model) did not come at the cost of poor function on the second task as the accuracy on this activity remained at 100% throughout the manipulation. As noted in Section 1, the potential neural plasticity of adult patients has been strikingly demonstrated through the fact that in LGG large swathes of eloquent tissue can be removed without causing an impairment (Desmurget et al., 2007). The same was true in this LGG simulation: after 1.4 million trials (as shown in Fig. 3), the affected hidden layer could be completely removed without performance falling below 90%, demonstrating that the unlesioned hidden layer was now mediating both tasks successfully.

4. Discussion

As well as providing a critical “engineering” test for theoretical assumptions, neurocomputational models also allow us to explore the underlying mechanisms that lead to these key clinical/behavioural outcomes. Specifically the LGG simulations prompt the question: are all units in the unlesioned hidden layer contributing equally to both tasks or has this layer become subdivided into more or less independent groups of units? The answer is emphatically the latter. Fig. 2 depicts the summed input from Input 2 → Hidden 2 and Input 1 → Hidden 2 before and after the decay regime. As can clearly be seen, approximately 50% of the units in the unlesioned hidden layer (Hidden 2) become unresponsive to activity in the input layer (Input 2) that was originally the sole driver for their performance prior to damage; instead, they are dominated by inputs from the lesioned subnetwork. Intriguingly, then, despite the fact that these “latent” cross-connections provided no useful input to either task (and could be lesioned without any effect) prior to damage—they are the driver to the plasticity-related changes in the model following damage.

Finally, it is not clear to what degree the specific parameter settings used in these networks are responsible for the fit to the data. Thus, we ran two sets of simulations exploring the interaction between two of our central computational principles; specifically the degree of entrenchment (analogous to age at lesion onset) and severity of the lesion. In the case of stroke, the severity of the lesion was parameterized as the number of hidden units deleted from the affected layer, while in the tumour simulations variability in severity was instantiated in the strength of the weight decay parameter.

5. Experiment 2

5.1. Methods

While most of the simulation parameters remained the same in these simulations, we varied entrenchment strength and lesion severity, under the hypothesis that these two factors would enter into a trading relationship, such that lower entrenchment (“younger” models) would lead to greater neural resilience. In the stroke simulation, the severity of the lesion was varied from 50 to 100% of the units in the affected hidden layer, in increments of 10%. The entrenchment varied from a cost strength of 2.25 to 5.25 in increments of 0.5. For the tumour simulations, the decay parameter varied between 0.0000001 and 0.000012475, in equal increments; the entrenchment manipulation was the same as in the stroke simulations.

5.2. Results

As can be seen in Fig. 4, the parametric manipulation of entrenchment and lesion severity yielded the predicted spread in post-lesion recovery. Specifically, a lower degree of entrenchment allowed for recovery from even the most severe of lesions, while high entrenchment led to significant deficits even in comparatively milder lesions. This was true both in the case of stroke, in which a percentage of the hidden units in one layer were lesioned, as well as in the progressive decay/tumour simulations, in which the relevant manipulation was the speed with which connections were reduced to 0, corresponding to behavioural outcomes observed in low- vs. high-grade tumour (Thiel et al., 2006).

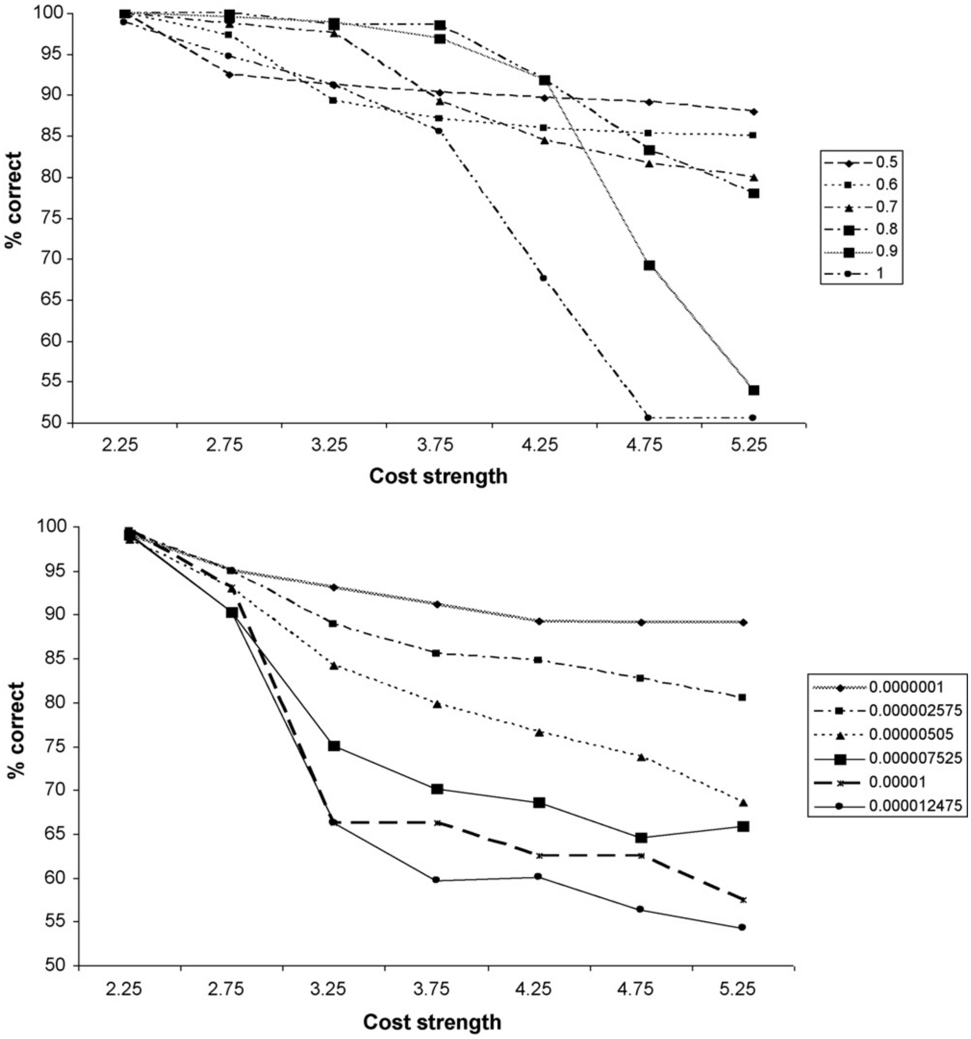

Fig. 4. Parametric variation of acute damage severity (simulated stroke) vs. speed of decay (simulated tumour) at different levels of cost strength (entrenchment/age). *Top panel*: Parametric manipulation of lesion extent and entrenchment. Each line in graph corresponds to a percentage of hidden units removed in simulated stroke (50–100%). *Bottom panel*: Manipulation of speed of decay (corresponding to low- vs. high-grade tumours) and entrenchment. Each line in graph corresponds to a value of weight decay.

6. General discussion

In this paper we have introduced a neurocomputational framework that provides new insight into the contrasting recovery profiles observed in acute vs. slow-growing lesions—and thus provides a potential solution to the paradox of modularity vs. equipotentiality in the adult brain. Although the neurocomputational model was based on a network of equipotential units and connections, the initial phase of development resulted in a modular organization with two emergent subnetworks specialized for two behavioural tasks. As would be expected from patients with stroke, when one of these subnetworks was damaged acutely then a behavioural dissociation resulted (task 1 impaired, task 2 unaffected), after a period of only partial recovery. Strikingly, a very different pattern emerged in the simulation of LGG. Slow decay of the links within the same subnetwork led to minimal performance decline, as reported in the patient literature (Duffau, 2005, 2006). Furthermore, at the end of the decay regime, the entire affected hidden layer could be removed with no effect on performance—which closely matches the lack of major impairment from LGG resection.

There are other important differences in the neural pathology associated with stroke vs. that observed in LGG. The abrupt loss of blood flow to regions of cortex that occurs in both ischemic and hemorrhagic stroke occasions relatively rapid neuronal death. In contrast, the slow development of LGG (which to a first approximation can be thought of as a pathological overproduction of well-differentiated glial cells) initially spares neuronal tissue. In fact, evidence from fMRI and direct electrical stimulation indicates that essential function can persist within the tumour for years; it is on this basis that a multi-stage surgical approach to these lesions has recently been proposed (Gil Robles, Gatignol, Lehericy, & Duffau, 2008). In both cases, however, the end result is loss of a significant amount of cortical tissue with very contrasting neuropsychological outcomes.

Our neurocomputational model highlights three key principles: patterns of connectivity; entrenchment of knowledge; and the time course of damage. While many computational models have utilized full connectivity, this is neurally implausible because full connectivity in the brain would require an unfeasibly large cranium (on the order of a sphere with radius 10 km: Nelson & Bower, 1990). Previous simulations have demonstrated that reduced connectivity shifts the functioning of units away from equipotentiality towards a form of graded quasi-modularity (Plaut, 2002). In the present neurocomputational model, a relatively strong form of quasi-modularity emerged by reducing cross-subnetwork connectivity. This suggests that neural systems do not need to be directly or even genetically pre-programmed for quasi-modularity of function to emerge but instead reflected their pattern of connectivity. In line with this notion, when connectivity is surgically re-engineered then the functioning of the newly connected cortical regions reflect the characteristics of the new domain and not the old (Sur, Garraghty, & Roe, 1988).

The approach to lesion-induced impairment and post-lesion plasticity in this model differs importantly from most previous modelling work. Typically, the majority of computational studies of impaired performance have simulated patient data by training a model to optimal performance and then damaging it and/or adding noise. The present study takes a somewhat different approach and, in doing so, builds upon a small set of previous investigations. Hinton and Sejnowski (1986) were probably the first to explore the impact of retraining after lesions to a computational model of reading. Marchman (1993) explored the impact of retraining after damage at different time points during development in a model of past tense verb processing. Plaut (1996) used a model of deep dyslexia (mapping orthography to phonology via semantic representations) and, like Hinton and Sejnowski, demonstrated that simulated recovery/therapy was faster than original development. Additionally, Plaut showed that the retraining outcome was better if non-prototypical concepts were used in the retraining set. Finally, more recently simulations have shown that the combination of damage and partial recovery can produce neuropsychological dissociations in the absence of in-built modularity. This is because plasticity-related recovery can alter the relative contributions of the remaining computational resources to maximise overall performance, a process that sometimes benefits some but not all items (Welbourne & Lambon Ralph, 2005). The present simulations build on these previous computational explorations and, perhaps most importantly, show how three key computational principals influence the degree and form of recovery/plasticity exhibited by such models (see below).

A key aspect of the original learning in this model and the subsequent recovery is experience-dependent plasticity. Development and adult learning in this type of neurocomputational model reflects continued exposure to a learning environment. This has been used extensively to simulate various aspects of normal and abnormal development in children (Munakata & McClelland, 2003). Classically, development is sometimes viewed as a form of learning fuel which is present in childhood but is gradually used up as we reach adulthood, at which point further large-scale restructuring is impossible. These neurocomputational models and other neuroscientific research (see below) suggest that a better analogy might be a chemical reaction: many chemical reactions do not really “finish” but rather slow down and stop when equilibrium is reached. However, they can be reactivated by a change in the chemical environment that disturbs this equilibrium. These computational models, and perhaps the brain, can be considered in a similar fashion: learning reflects a process by which a balance is found between the behavioural challenges of the environment with the computational resources and connectivity at hand. The apparent stability or even rigidity of the adult system represents the equilibrium point. When this is disturbed by brain damage, the remaining resources (including latent connectivity) will no longer be optimized for performance and thus a new equilibrium is found through plasticity-related changes.

The most critical aspect of the present neurocomputational model is the time-course of damage. Only partial, plasticity-related recovery followed after acute damage, yet full-blown, effective function can be maintained by gradual damage. This reflects an interaction between the speed of damage and the entrenchment of knowledge (the bias to maintain existing knowledge structures captured by the cost function applied to the model's output units—see Section 2.1). When the model is damaged acutely on a large scale (to simulate stroke), the knowledge structure of the model is decimated. This means that there is a large disparity between target behaviour and actual performance. As a consequence, while plasticity-related changes can reduce this disparity (and simultaneously maintain performance on the unaffected task) it can never do so optimally. Although large-scale performance differences are also present in the developing undamaged model, effective development proceeds because both tasks are at the same stage. Slow damage is much less demanding on plasticity-related changes because after each small episode of damage only small scale adjustments are required to maintain optimal performance on both tasks. Over time, these gradual adjustments lead to an effective reorganization of the remaining undamaged network.

These principles of adult neural plasticity are reified in the effects of the age-related entrenchment factor on relearning following damage to the model. From a computational standpoint, the entrenchment factor provides an important and novel way to simulate impairment within a PDP framework beyond the simple limitation of resources. As described above, simulation of permanent neural damage in single network frameworks requires the removal of enough units or connections that the model is mathematically incapable of performing a given task—equivalent to an assertion that the lack of recovery in stroke results from a lack of resources. However, the contrasting recovery profiles explored in this paper suggest that this cannot be the whole explanation. If correct, then there must be some additional computational mechanism which limits the potential recovery in the stroke case, but not in the LGG case. We believe that our entrenchment factor is a good candidate for this role.

Beyond these practical computational considerations, the entrenchment factor instantiates what we view as a basic tradeoff between the high adaptability characteristic of (and essential for) successful neural development and the representational rigidity of the adult state evinced, for example, by the typical inability of adults to acquire fluency in a second language. Early in the development of the model, when the strength of the function is low, the effect of entrenchment is very low. As its strength increases, it makes wholesale functional changes much less likely than they would be if the brain maintained a constant level of plasticity throughout life. The developing system needs to be highly receptive to the statistical structure of the environment in order to structure itself in accordance with outside demands. However, it is optimal for this receptivity/plasticity to gradually reduce in favour of increasing stability in the adult state, since the environment does not typically change in a large-scale fashion throughout the lifespan.

The effects of entrenchment can be understood through the contrast of two types of learning in the adult state. For instance, adult native Japanese speakers display a remarkable inability to discriminate the English phonemes /r/ and /l/, even after extensive training on this distinction (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999). In this situation, the phonemic distinction requires a restructuring of the existing representations in an intact system. Such phenomena are easily accounted for in terms of native language attractors rendering a rather large acoustic difference less perceptible for these subjects at the phonemic level—this can be viewed as an entrenchment effect. In cases where the attractors are insufficiently deep or powerful, as is the case early in development or when the new phoneme does not fall between two existing attractors, learning is typically successful. The second type of learning involves reestablishment of representations destroyed by lesion; for example, in the case of extensive damage to the left IFG. Here, interference from existing representations cannot explain the lack of recovery because these representations were supported by the damaged cortex and thus this knowledge is lost in patients. In this situation, therefore, relearning must occur *de novo*. However, the many reports of chronic aphasia in the literature, as well as the results of our simulations, suggest that the adult brain does not successfully accomplish this relearning in the face of acute insult but does so if the damage is gradual and extended over time.

Anatomically, this difference is reflected in patterns of reorganization observed in stroke and tumour. In stroke, successful recovery is typically mediated by perilesional tissue while activations in areas not associated with the function premorbidly often reflect incomplete recovery. Thus, in the case of stroke-induced nonfluent aphasia, successful recovery is associated with compensatory activity in the left temporal lobe. Activity in the contralateral IFG (which is not typically activated during language processing in healthy right handed subjects) is thought to result from a loss of transcalosal inhibition and has been claimed to be unrelated to recovery or even to be maladaptive (see Price and Crinion, 2005 for a review). In contrast, successful language reorganization in LGG patients involves a number of atypical structures and pathways. Thiel et al. (2001) imaged 61 left hemisphere tumour patients while performing a verb-generation task. While controls mainly activated typical left hemisphere language pathways, tumour patients recruited a number of areas not normally observed in language tasks. In both the ipsilesional and contralesional hemispheres, activations were consistently detected in atypical frontal areas. Additionally, Thiel et al. (2005) induced transient language disturbance after rTMS to right IFG in five left-handed, left hemisphere tumour patients, thus confirming a shift of essential language function to the contralateral hemisphere quite unlike the pattern observed in stroke aphasia. These studies also support the key principle of time-course-of-damage in our neurocomputational model: Thiel et al. (2006) demonstrated that this large-scale reorganization does not follow in case of fast tumour.

Further evidence for experience-dependent plasticity can be found in other parts of the neuroscience literature: for example, Dancause et al. (2005) observed the establishment of novel connections in squirrel monkey motor cortex following experimentally induced ischemic infarct in M1. Additionally, Bridge, Thomas, Jbabdi, and Cowey (2008) described findings from diffusion-weighted imaging in the 'blindsight' subject GY. While both GY and controls exhibit an ipsilateral connection from LGN to MT+/V5, an additional contralateral pathway was demonstrated in GY, suggesting a change in connectivity as a result of this lesion. The sometimes negative impact of knowledge entrenchment likewise has correlates in basic neuroscience: in a series of highly influential studies Knudsen and his colleagues have studied the ability of barn owls to adjust to new mappings between visual and auditory input induced by prism-shift goggles (Brainard & Knudsen, 1998; Linkenhoker & Knudsen, 2002). While juvenile barn owls accommodate rather large disparities in this mapping, adult barn owls do not adapt to such large shifts. However, very much like the slow-paced adjustments of the LGG simulation, Linkenhoker and Knudsen (2002) demonstrated that adult barn owls could in fact achieve re-mappings almost as extreme when the owls were fitted with prisms which progressively shifted visual input by a few degrees at a time.

Taken together, these results suggest a middle ground between a strictly modular view of the brain and the Lashleyan ideal of cortical equipotentiality. While it is likely that we have yet to observe the full extent of adult neural plasticity, we hope here to have identified a small set of computational principles that provide a framework for understanding the sequelae of brain lesion. Neural architecture is shaped by a combination of its current pattern of connectivity, the statistical structure of the environment and plasticity-related changes that continue throughout the lifespan. It is only through a full understanding of these neurocomputational principles that effective interventions and therapies will be achieved for patients with neurological damage. This study is intended as a first step in that important journey.

Acknowledgement

This work was supported by a grant from the Gatsby Foundation (GAT2831).

Appendix

While the results presented in this paper suggest that the effects of knowledge entrenchment are effectively simulated through use of the logistic cost function, the question remains as to why. That is, what difference between the two models allows the LGG model to maintain a high level of performance while the stroke model, with the same architecture and resources, does not. An answer to this question lies in the dynamics of the entrenchment factor, and how it interacts with the errors induced by the lesion.

To better understand this interaction, we must first examine how PDP networks learn in the typical (undamaged) case, focusing the factors that govern how weights entering the output layer are changed on the basis of mismatch between their target (t) and their actual output (o). On each trial, a pattern is applied to the units of the input layer—these values are known as 'clamps' and represent the output of the input units. Each unit in the following (hidden) layer receives a weighted sum of these clamps (its 'input'), equal to:

$$i(j) = \sum w_{ji} \times o_i$$

for all sending units i, where w_{ji} is the weight to unit j from unit i, and o_i is the output of the sending unit. To calculate the output of this unit, the input is passed through the activation (or 'squashing') function, in this case the logistic activation function, to yield the output:

$$o_i = \frac{1}{1 + e^{-i(j)}}$$

Thus, in cases of large positive input the output will be very close to 1, while in the case of strong negative input the output will be close to 0. The activations of the output units are calculated in the same way, as a weighted sum of the outputs of the hidden units. For illustrative purposes, let us assume that the target value for some output unit j is 1.0, while the actual output on this trial is 0.7. According to the delta rule, we change the weight entering j from some hidden unit i according to

$$\Delta w(ji) = \eta \times (t_j - o_j) \times o_i$$

where η is a constant called the learning rate, in our example 0.01. Thus, for our hypothetical unit, the weight w_{ji} will be changed by $0.01 \times 0.3 \times o_i$, and this weight change will ensure that the next time this training pattern is presented the output of this unit will be closer to the target, all other things being equal.

This is the manner in which the models in this paper were trained up until the simulation of a lesion. In the case of the LGG simulation, on each trial every weight going into and out of the affected hidden layer was multiplied by a small decay factor (in Experiment 1, 1.0×10^{-6}), and the absolute value of this product was subtracted from the weight. The effect of this is, on average, to create a small amount of error on the next presentation of the same pattern. For instance, if the target for a given output unit is 1.0, the actual output after decay will likely be lower than this, as an example let us say the output is 0.8. While this is still correct, in the sense of being on the right side of 0.5, it also yields an error signal because it does not exactly match the target. However, due to the addition of the entrenchment factor, the effect of this error signal on the weight change is no longer straightforward. As stated in the main text, entrenchment is instantiated as a logistic cost function, whose inverse is:

$$\text{Cost} = \frac{(\ln(o_j) - \ln(1 - o_j)) \times C_s}{\ln(C_p)}$$

Since the cost peak remains constant at 0.5, we can ignore its effect and focus on the remaining terms. Clearly, $(\ln(o_j) - \ln(1 - o_j))$, where o_j represents the activation of the output unit, reaches extreme values in cases where the output is close to 1 or 0. This value is then multiplied by the cost strength (a constant scaling factor), and this product is added to the term $(t_j - o_j)$.

Since the learning rate on the affected links of our LGG simulation is set to 0, we need only focus on the sparse cross-connections. Because the error in the LGG case is typically a small deflection away from the target, the contributions from the error signal $(t_j - o_j)$ and the entrenchment factor are in the same direction, and thus both serve to reduce the error on future presentations of a given trial. However, in the case of acute damage such as stroke, the errors are often of large magnitude, and in this case the entrenchment factor can overwhelm the error signal. For instance, in a case where a stroke leads to an output of 0.2 on a unit whose target is 1.0, the error signal will encourage a weight change that would lead to a larger output on the next trial (i.e., closer to 1), while the entrenchment factor will favour a weight change that leads to an output of 0. Depending on the cost strength (which increases throughout the 'development') phase on the model, the resultant weight change will either be adaptive (in the case of low cost strength) or maladaptive (in the case of high cost strength). In both cases, the outcome represents the effects of the tradeoff between high neural plasticity and entrenchment of stored knowledge structures.

References

Bao, S. W., Chang, E. F., Davis, J. D., Gobeske, K. T., & Merzenich, M. M. (2003). Progressive degradation and subsequent refinement of acoustic representations in the adult auditory cortex. *Journal of Neuroscience*, *23*(34), 10765–10775.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex*, *14*(10), 1088–1099.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English vertical bar r vertical bar and vertical bar l vertical bar: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985.

Brainard, M. S., & Knudsen, E. I. (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *Journal of Neuroscience*, *18*(10), 3929–3942.

Bridge, H., Thomas, O., Jbabdi, S., & Cowey, A. (2008). Changes in connectivity after visual cortical brain damage underlie altered visual function. *Brain*, *131*, 1433–1444.

Curtiss, S., de Bode, S., & Mathern, G. W. (2001). Spoken language outcomes after hemispherectomy: Factoring in etiology. *Brain and Language*, *79*(3), 379–396.

Dancause, N., Barbay, S., Frost, S. B., Plautz, E. J., Chen, D. F., Zoubina, E. V., et al. (2005). Extensive cortical rewiring after brain injury. *Journal of Neuroscience*, *25*(44), 10167–10179.

Desmurget, M., Bonnetblanc, F., & Duffau, H. (2007). Contrasting acute and slow-growing lesions: A new door to brain plasticity. *Brain*, *130*, 898–914.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training—Newly honed juggling skills show up as a transient feature on a brain-imaging scan. *Nature*, *427*(6972), 311–312.

Duffau, H. (2005). Lessons from brain mapping in surgery for low-grade glioma: Insights into associations between tumour and brain plasticity. *Lancet Neurology*, *4*(8), 476–486.

Duffau, H. (2006). New concepts in surgery of WHO grade II gliomas: Functional brain mapping, connectionism and plasticity—A review. *Journal of Neuro-Oncology*, *79*(1), 77–115.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573.

Gil Robles, S., Gatignol, P., Lehericy, S., & Duffau, H. (2008). Long-term brain plasticity allowing a multistage surgical approach to World Health Organization Grade II gliomas in eloquent areas. *Journal of Neurosurgery*, *109*(4), 615–624.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In J. L. McClelland, & D. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 282–317). Cambridge: MIT Press.

Liegeois, F., Connelly, A., Baldeweg, T., & Vargha-Khadem, F. (2008). Speaking with a single cerebral hemisphere: fMRI language organization after hemispherectomy in childhood. *Brain and Language*, *106*(3), 195–203.

Linkenhoker, B. A., & Knudsen, E. I. (2002). Incremental training increases the plasticity of the auditory space map in adult barn owls. *Nature*, *419*(6904), 293–296.

Mandonnet, E., Delattre, J. Y., Tanguy, M. L., Swanson, K. R., Carpentier, A. F., Duffau, H., et al. (2003). Continuous growth of mean tumor diameter in a subset of grade II gliomas. *Annals of Neurology*, *53*(4), 524–528.

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, *5*(2), 215–234.

Muller, R. A., Rothermel, R. D., Behen, M. E., Muzik, O., Chakraborty, P. K., & Chugani, H. T. (1999). Language organization in patients with early and late left-hemisphere lesion: A PET study. *Neuropsychologia*, *37*(5), 545–557.

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6*(4), 413–429.

Nelson, M. E., & Bower, J. M. (1990). Brain maps and parallel computers. *Trends in Neurosciences*, *13*(10), 403–408.

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language*, *52*(1), 25–82.

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, *19*(7), 603–639.

Price, C. J., & Crinion, J. (2005). The latest on functional imaging studies of aphasic stroke. *Current Opinion in Neurology*, *18*(4), 429–434.

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421.

Rao, S. M., Binder, J. R., Hammeke, T. A., Bandettini, P. A., Bobholz, J. A., Frost, J. A., et al. (1995). Somatotopic mapping of the human primary motor cortex with functional magnetic-resonance-imaging. *Neurology*, *45*(5), 919–924.

Sur, M., Garraghty, P. E., & Roe, A. W. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science*, *242*(4884), 1437–1441.

Thiel, A., Habedank, B., Herholz, K., Kessler, J., Winhuisen, L., Haupt, W. F., et al. (2006). From the left to the right: How the brain compensates progressive loss of language function. *Brain and Language*, *98*(1), 57–65.

Thiel, A., Habedank, B., Winhuisen, L., Herholz, K., Kessler, J., Haupt, W. F., et al. (2005). Essential language function of the right hemisphere in brain tumor patients. *Annals of Neurology*, *57*(1), 128–131.

Thiel, A., Herholz, K., Koyuncu, A., Ghaemi, M., Kracht, L. W., Habedank, B., et al. (2001). Plasticity of language networks in patients with brain tumors: A positron emission tomography activation study. *Annals of Neurology*, *50*(5), 620–629.

Wade, D. T., Hewer, R. L., David, R. M., & Enderby, P. M. (1986). Aphasia after stroke-natural-history and associated deficits. *Journal of Neurology, Neurosurgery and Psychiatry*, *49*(1), 11–16.

Welbourne, S. R., & Lambon Ralph, M. A. (2005). Using computational, parallel distributed processing networks to model rehabilitation in patients with acquired dyslexia: An initial investigation. *Aphasiology*, *19*(9), 789–806.

Welbourne, S. R., & Lambon Ralph, M. A. (2007). Using parallel distributed processing models to simulate phonological dyslexia: The key role of plasticity-related recovery. *Journal of Cognitive Neuroscience*, *19*(7), 1125–1139.

Young, M. P., Scannell, J. W., & Burns, G. (1995). *The analysis of cortical connectivity*. Heidelberg: Springer-Verlag.