A unified neurocomputational bilateral model of spoken language production in healthy participants and recovery in post-stroke aphasia

Ya-Ning CHANG[1]*, Matthew A. LAMBON RALPH[1]*

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

Corresponding to:
MRC Cognition and Brain Sciences Unit
University of Cambridge
15 Chaucer Road
Cambridge
CB2 7EF
UK

*Email: yaning.chang@mrc-cbu.cam.ac.uk
matt.lambon-ralph@mrc-cbu.cam.ac.uk

ORCiD: Ya-Ning Chang (0000-0001-5248-0761)
Matthew A. Lambon Ralph (0000-0001-5907-2488)

Classification: Psychological and Cognitive Sciences;

Keywords: stroke aphasia, language lateralisation, language recovery, bilateral language processing, neurocomputational modelling

Author Contributions: Y.C. and M.A.L.R designed and performed research; Y.C. and M.A.L.R analysed data; and Y.C. and M.A.L.R wrote the paper.

This PDF file includes:
Main Text
Figures 1 to 5
Table 1
Supplementary

Abstract

Language is a critical human ability. When impaired, it has significant impacts on everyday life. Therefore, understanding of the processes underlying normal, impaired and recovered language performance has been a long-standing goal for both cognitive and clinical neuroscience. Vibrant studies of healthy and impaired language have generated many verbally-described hypotheses about language lateralisation and recovery. However, they have not been considered within a single, unified and implemented computational framework, and the literatures on healthy participants and patients are largely separated. These investigations also span different types of data, including behavioural results and fMRI brain activations, which augment the challenge for any unified theory. Consequently, many key issues, apparent contradictions and puzzles remain to be solved. Here, we developed a neurocomputational, bilateral pathway model of spoken language production, designed to provide a unified framework to assimilate different types of data from healthy participants and aphasic patients. The model encapsulates various key computational principles (differential computational resources, emergent division of labour across pathways, experience-dependent plasticity-related recovery) and provides an explanation for the bilateral yet asymmetric lateralisation of language in healthy participants, chronic aphasia after left rather than right hemisphere lesions, and the basis of partial recovery of function in patients. Also, the model provides a formal basis for understanding the relationship between behavioural performance and brain activation. Overall, the unified model is consistent with the degeneracy and variable displacement theories of language recovery, and adds computational insights to these hypotheses regarding the neural machinery underlying language processing and plasticity-related recovery following damage.

Significance Statement

Studies of healthy and impaired language are thriving and generate many verbally-described hypotheses. Whilst these verbal descriptions have advanced our understanding of language processing, some explanations seem to be mutually incompatible. It remains unclear exactly how different descriptions work mechanistically. Here, we constructed a neurocomputational bilateral model of spoken language production to simulate a range of phenomena in healthy participants and aphasic patients simultaneously including partial lateralisation in healthy participants, left but not right stroke aphasia, and the role of the hemispheres in plasticity-dependent recovery. The resulting patterns demonstrate how individual and seemly contradictory findings can be assimilated within a single framework. To our knowledge, this provides the first coherent mechanistic account of language lateralisation and recovery from post-stroke aphasia.

Introduction

Language is a key human ability and when impaired (e.g., after stroke or neurodegeneration), patients are left with significant disability in their professional and everyday lives. These language impairments are common - around one-third of the 10 million+ patients in the acute phase post stroke (1). Both studies of healthy and impaired language have a long history, and these vibrant literatures have generated many verbally described hypotheses, including notions around healthy language, impaired language and how it might partially recover after brain damage. In particular, the long-standing literature on language impairment in aphasia dates back to seminal 19th century studies (2-4). While these verbally described hypotheses have advanced our understanding of language processing both theoretically and clinically, it is not clear how some of the descriptions work mechanistically and they can be mutually incompatible. For instance, some notions propose good aphasia recovery only results from language returning to the left hemisphere (5-9) while others report that the recovered language performance is positively correlated with activation in the right hemisphere (10-12). As a recent review by Stefaniak et al. (13) noted, the current situation is confusing because there are many individual findings, different types of data (e.g., patients' language performance vs. fMRI activations) yet no unified mechanistic account. There is a pressing need to have an implemented neurocomputational model which can provide: (a) a unified framework in which findings from healthy participants and aphasic patients can be assimilated; (b) a computationally-instantiated framework to formalise and test verbally-described hypotheses; and (c) a framework that can bridge between different types of cognitive neuroscience data including language behaviour, lesion locations and task-related fMRI. This was the overarching aim of the current study. The puzzles and targets are set out briefly below.

Lateralisation assumptions from fMRI in healthy participants versus chronic aphasic patients

The first issue concerns lateralisation assumptions from healthy and impaired language. The very strongly held view that language is a left hemisphere function primarily arises from the long-standing neuropsychology literature showing that chronic aphasia is associated with left hemisphere damage but is not generally associated with right hemisphere damage (14-16). However, patient data are perhaps more graded than often portrayed. Recent evidence has shown that right hemisphere lesions can generate language problems especially in the early phase and some mild remaining deficits can be measured in chronic cases (17). Moreover, several transcranial magnetic stimulation (TMS) studies of semantics (18-21) or phonology (22), and patient studies of semantics (23-25) indicate that left and right areas contribute to healthy language performance, and in some cases bilateral damage is required to show more substantial deficits.

In contrast, the rise of functional neuroimaging in healthy participants has shown many language tasks such as repetition, picture naming, comprehension and production are bilaterally supported (26-31). Although the activation patterns are often leftward asymmetric, the degree of asymmetry largely depends on the nature of the tasks with a subset showing stronger forms of asymmetric bias. For instance, propositional speech production tasks are more left lateralised, and involve greater activation in the left inferior frontal gyrus, whereas nonpropositional speech production tasks (e.g., counting) involve more bilateral activations (32-34). When considering findings from both chronic aphasic patients and healthy participants, it appears difficult to reconcile the seemingly contradictory findings: how can the language network be strongly left lateralised in patients but be bilateral, albeit asymmetric, in healthy participants?

We propose that these results could reflect the outcome of an intrinsically bilateral but asymmetric language network for speech production. Functional asymmetry could follow from hemispheric asymmetry in language areas (35-41). For instance, using a combined fMRI and diffusion tensor imaging, Vernooij et al (41) demonstrated a significant correlation between the

functional hemispheric lateralisation and the relative fibre density asymmetry of the arcuate fasciculus in right-handed subjects. Although the sample size in Vernooij et al. was small, similar findings were also reported in Ocklenburg et al. (40) based on a relatively larger sample size. Thus, within the language network, the majority of right-handed healthy participants show leftward asymmetry of brain volumes and arcuate fasciculus (35, 39, 42), suggesting that more of the computational resources are in the left than the right hemisphere. Such resource imbalance should generate a bilateral yet asymmetric lateralisation in simulated Blood Oxygen Level Dependent (BOLD) and also greater likelihood of chronic impairment after left than right damage. The latter may reflect a combination of the premorbid division of labour for left over right in healthy language but also the potential for plasticity-related recovery post damage. This has been explored for specific language tasks in past computational work by re-exposing the damaged model to its learning environment, generating plasticity-related recovery via "retuning" of the remaining computational resources (43, 44). A straightforward hypothesis, from these earlier models, is that the potential for such recovery reflects the amount of computational resource available. Accordingly, a smaller right hemisphere contribution to language will also mean less potential for picking up additional language work post damage.

The computational bases of language recovery

The second critical issue concerns the computational bases of language recovery. A recent review (13) considered two mechanisms that may underpin language recovery: *degeneracy* and *variable neuro-displacement*. Degeneracy suggests that, like other biological systems, brain function might be multiply coded across different regions and/or pathways resulting in a partially resilient system. Borrowing from engineering, variable neuro-displacement suggests that normal brain function may be engineered to be resilient to variations in task demand and also to minimise energy expenditure, given that the brain is a very metabolically expensive organ. Accordingly, brain

functions are implemented across neural networks with additional capacity in them that is dynamically titrated according to ongoing task performance. Both mechanisms provide the bases for some degree of resilience to damage and potential for recovery of function following damage via a permanent reformulation of the remaining multiple codes (degeneracy) and/or upregulation of systems (variable neuro-displacement). Previous computational studies (43, 44) of plasticity-related recovery have provided some support for these principles by demonstrating that re-exposing a damaged model to its learning environment leads to two types of experience-dependent learning, depending on remaining resources in the system. If the model only has a single pathway to perform the task, re-learning can retune and upregulate the contribution of 'perilesional' units and weights. Secondly, if there are multiple routes that support the task, re-learning can also shift the division of labour between different pathways in the system, in addition to potential upregulation of perilesional units. The potential for recovery-related changes is likely to be determined by the relative resources available in different pathways and their engagement in the task prior to damage (i.e., premorbid status). Though interesting, these mechanistic hypotheses about language recovery need to be explored more formally within an implemented computational model and preferably one that can simulate healthy and impaired language, as well as generate the different measures used to assess recovery of function, such as language performance and fMRI activations.

Theories of aphasia recovery

The long-standing literature on language recovery in post-stroke aphasia has generated a very large number of hypotheses. However, most hypotheses are verbally described or are verbal descriptions of observed phenomena (5). Two high-profile well-rehearsed notions can be considered as examples. First, upregulated activation in perilesional and contralesional areas has been associated with recovered performance in post-stroke aphasia (5, 45-50). Upregulated activation could be viewed as an example of variable neuro-displacement. That is, the broader activation clusters

observed in healthy fMRI data might be upregulated permanently to support recovered function in patients when tasks are made harder or the statistical threshold dropped. For example, van Oers et al. (49) showed that recovery on picture naming in post-stroke aphasic patients was associated with activation in the left inferior frontal gyrus (IFG) while recovery on more cognitively demanding task (e.g., the Token Test) was associated with upregulated contralesional activation in the right IFG in addition to the left IFG. There is also parallel evidence from combined TMS-fMRI studies in healthy participants that inhibition of the left anterior temporal lobe (ATL) upregulates activation in the right ATL to support semantic tasks (18, 19).

A second well-rehearsed notion about aphasia recovery is the right hemisphere hypothesis (RHH). Despite being a commonly repeated hypothesis dating over a century, as far as we are aware, there are no implemented bilateral language models in which the notion can be formally evaluated. RHH can be considered as an example of variable neuro-displacement or degeneracy mechanism (13). Numerous fMRI and positron emission tomography (PET) studies have demonstrated that patients with chronic damage in the left hemisphere recruit the right hemisphere during language tasks with greater right hemispheric activation in patients than in healthy participants (8, 46, 51). These findings have been interpreted in terms of a right hemisphere juvenile "back-up" language system, which is weaker and error-prone. Normally suppressed by the dominant left hemisphere system, it can be released to provide some function after significant left hemisphere damage. The picture is made more confusing because the hypotheses and data in relation to the RHH are contradictory. Some notions suggest that aphasia recovery is supported by this right hemisphere system; when aphasic patients have a second stroke in the right hemisphere, their language performance generally becomes worse (2, 52). There is also evidence that language performance is correlated with activation in the right hemisphere (10-12). In contrast, the 'regional hierarchy framework' proposes that right hemisphere activation is maladaptive and good recovery only results from language returning to the left (5-9). According to a seminal study of post-stroke aphasia by

Saur et al. (46), left hemisphere activation for auditory comprehension greatly decreased a few days after stroke, was followed by increased bilateral activation with a significantly upregulated peak in the right hemisphere two weeks after stoke, and then the peak activation shifted back to the left hemisphere in the chronic phase. Given that the patients in Saur et al.'s study had very mild aphasia and showed excellent recovery of function, the finding seems to suggest that right hemisphere activation is associated with initial recovery, yet better long-term recovery may require activations to shift back to a more typical left lateralised pattern. However, it remains unclear what mechanisms underlie the changes in brain activity and what the longitudinal patterns are for moderate and severe aphasia.

These RHHs have, in turn, inspired interventions with opposing aims: either promoting right hemisphere engagement (53) or trying to suppress it in favour of left hemisphere involvement using TMS or transcranial direct current stimulation (tDCS) (54-57). Without a better understanding of underlying mechanisms and a formal implemented model, various foundational issues remain. These include: how a right hemisphere system can develop if it is suppressed by the left hemisphere; how the two systems might interact; whether the results of negative associations between right hemisphere activation and language are simply a reflection of behavioural severity and lesion size, as mild aphasia is associated with small lesions which leaves more of the left hemisphere intact and able to be activated. Our working assumption is that there is an intrinsically bilateral, albeit asymmetrically-provisioned single functional network. That is, the left and right hemispheres both contribute to speech production with differential contributions arising from the effects of imbalanced resources across the hemispheres. An implemented model would permit a proper investigation of how this division of labour might shift and under what conditions after brain damage.

Additionally, the hypothesis of the maladaptive right hemisphere activation within the regional hierarchy framework supposes that the two hemispheres attempt to inhibit each other through transcallosal inhibition (6, 7, 9). There are several puzzles about this hypothesis including

(a) why the healthy brain might spend most, if not all, of its lifetime inhibiting regions from working (a biologically expensive implementation) and (b) how the less dominant system can even develop semi-useful representations if being persistently suppressed. We also note that to the best of our knowledge – outside of the motor system (58-60) – there are no demonstrations of transcallosal inhibitory connectivity. But with an implemented bilateral language model, we can explore the effect of including transcallosal connectivity on healthy and impaired function.

Multiple measures

The last issue concerns different types of data and measures. Classically, explorations of brain function relied on relating lesions/brain damage to the pattern of patients' performance (61-63). The advance of functional neuroimaging techniques has allowed healthy and damaged function to be explored, *in vivo*. A corollary is that we now have multiple measures to consider in parallel, including lesion location and size, behavioural language measures, observed activations as well as connectivity. To make progress, the field needs to begin to understand the relationship between observed behavioural performance and brain activation, at different degrees and locations of brain damage. It is tempting to assume that activated regions must be contributing to the observed patient performance but, like any form of functional neuroimaging, simply observing activation does not mean that the region is critically contributing to healthy or impaired performance (64). This may explain inconsistent findings in which activation in right hemispheric language areas is not always correlated with language performance in post-stroke aphasia (8, 10-12, 48, 65, 66). Indeed, different types of imaging analyses, such as multiple voxel pattern analysis (MVPA) (67), have started to be used to predict stroke patients' recovered performance. Representational similarity analysis (RSA) has been applied to investigate the brain activation patterns of perilesional and contralesional regions after stroke and their involvements in language processing. For instance, Fischer-Baum et al. (68) reported that, when a stroke patient with severe written language impairment was asked to perform a

naming detection task, the orthographic activation patterns in the right fusiform gyrus were more similar to stimulus patterns than in the left fusiform gyrus. Thus, it is critical that a computational model can be designed to accommodate multiple measures within a single framework so that the relative levels of activation across layers (akin to regions of the brain) can also be probed. This allows a formal exploration of the relationship between brain activations and contributions to the observed behavioural performance.

To summarise, the primary aim of this study was to develop a unified, bilateral pathway model of spoken language production that could assimilate findings in healthy participants and in post-stroke aphasia to resolve several puzzles in the literature. Specifically, we investigated four key issues: (a) how the system might show bilateral albeit asymmetric activation in healthy participants but a very strong lateralisation in post-stroke patients; (b) how activation patterns change dynamically across the hemispheres during recovery; (c) the relationship between multiple measures including recovered behavioural performance and brain activation; and (d) if there is an effect of transcallosal connectivity on healthy and impaired function.

Results

Hemispheric asymmetry and language lateralisation

The bilateral model of language processing was implemented as a simple recurrent network. The model consisted of two parallel pathways. The model was trained to perform a repetition task (see the Methods section for details). We investigated if the model could simulate language lateralisation that follows hemispheric asymmetry, similar to the correlation pattern between functional hemispheric asymmetry in parietotemporal regions and structural asymmetry in arcuate fasciculus during the spoken production task reported in Vernooij et al (41). Specifically, we varied the proportion of hidden units in the left versus the right pathways in the model (see Fig. 1A) to simulate the relative capacity of the two pathways (69) while the total number of hidden units

remained unchanged. The number of units for the two consecutive hidden layers in both the left and right was the same. In each of the five capacity conditions, twenty versions of the model were trained with different random initial weights. The same training procedure was applied to each condition. After training, the model was tested on both the word and nonword repetition tasks.

In the imaging studies, a laterality index is generally estimated using the BOLD signals in the left and right homologue language areas, where the right BOLD signals are subtracted from the left BOLD signals and then dividing the score by the sum of them (41, 70). In our model, two different measures can be used to compute the degree of lateralisation, one is *functional contribution* and the other one is *output unit activation*. Functional contribution is a measure of the relevant contribution from the left or right pathway to output activation (44, 71). Alternatively, output unit activation measures average unit activation at the output layer uniquely from either the left or right pathway. It has been used as a proxy of the fMRI BOLD signals in previous simulation work (72). In the present study, we computed laterality indices based on both measures. The positive score indicated the model showed a left lateralised pattern; conversely, the negative score indicated the model showed a right lateralised pattern. More details about the computations of functional contribution and output unit activation are reported in the Methods section. An asymmetry index for model capacity was calculated in a similar way to the laterality index, in which the number of hidden units in the right was subtracted from the number of hidden units in the left and then dividing by the total. Moreover, we also investigated average unit activation across the hidden layers along the left and right pathways separately in different conditions to test if more hidden units (i.e., more processing resource) would lead to higher activation on average.

Results are summarised in Fig. 1. All models performed well on word repetition and generalised to nonwords (Fig. 1B). There was a clear lexicality effect with the highest accuracy for high frequency words followed by low frequency words and then nonwords. Importantly, the performance level achieved by the model with differential capacities in the left and right was very

similar because the total resources were the same. These observations were confirmed by a repeated ANOVA result. There was a significant word type effect [HF vs. LF vs. nonwords: $F(2, 278) = 33.8$, $p < .001$], while both model type ($p = 0.44$) and its interaction with word type ($p = 0.18$) were not significant. This means that the model was able to exploit the computational resources flexibly to learn the task and to generalise. In contrast, the underlying processing did change. Figure 1C shows that more hidden units along the two processing pathways result in higher hidden unit activation. This suggests that the functional division of labour in the model was based not only on there being more units in the "dominant" processing pathway but they also resultantly worked harder on average. Critically, Figure 1D shows the model with more processing resources in the left pathway (i.e., a larger asymmetry index) produces a more left lateralised pattern. Laterality indices based on function contribution and output unit activation were both positively correlated with asymmetry index for model capacity, Pearson's $r = 0.929$, $p < .001$, and Pearson's $r = 0.917$, $p < .001$, respectively. The results are consistent with the function- structure pattern reported in Vernooij et al. (41). Together, these investigations provide computational evidence to link language lateralisation with imbalance processing resources in the left and right language areas, and reveal the consequent change in the functional division of labour underlying performance.

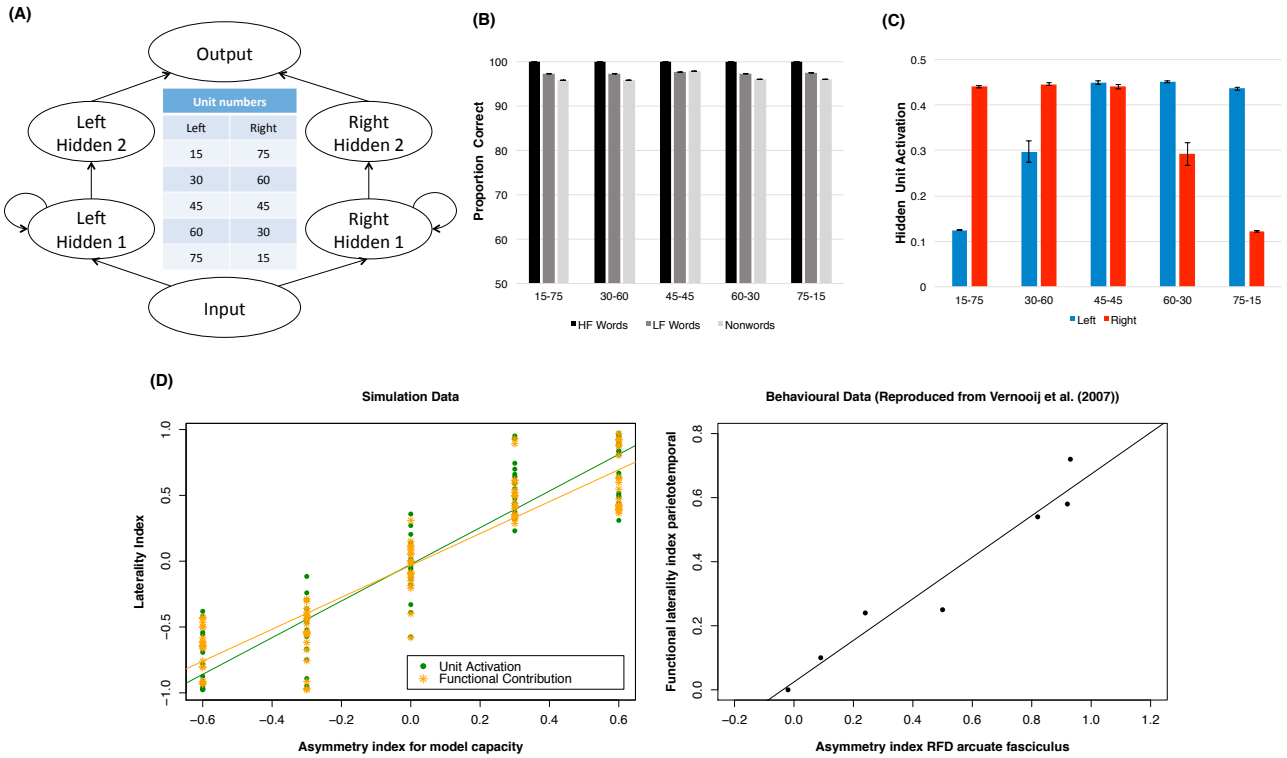

Figure 1. The model architecture, repetition performance, average hidden unit activation, and lateralisation patterns produced by the bilateral model with differential capacity in the left and right pathways. The behavioural data were reproduced from the spoken production task reported by Vernooij et al. (41). (A) The model with five different numbers of hidden units in the left and right pathways including 15-75, 30-60, 45-45, 60-30, and 75-15. The number of units in the hidden 1 layer and the hidden 2 layer was the same; (B) The repetition performance of the model on high frequency words, low frequency words and nonwords; (C) Hidden unit activation produced by the model across the hidden layers along the left and right pathways; (D) The lateralisation patterns based on functional contribution and output unit activation produced by the model and the behavioural data reproduced from Vernooij et al. (41). HF: high frequency; LF: low frequency. Error bar represents ±1 standard error.

Chronic aphasia as a consequence of left hemisphere stroke but not right hemisphere stroke

We next investigated whether damage to the left hidden layer in the model would be more likely to result in permanently impaired language performance (chronic aphasia) compared to damage to the right. As Gajardo-Vidal et al.'s (17) large case series study reported, approximately half of patients 49% (151/307) with left hemisphere stroke showed impaired repetition performance, whereas the incidence of impaired repetition performance for patients with right hemisphere stroke was about 5% (5/93). In the preceding section, we demonstrated that the model with more computational resources in the left pathway produced a bilateral, left-asymmetric activation pattern similar to fMRI brain activations observed in most healthy individuals during language production. Thus, we opted to use a model with an asymmetrical structure where the computational capacity in the left was twice as large as that in the right (60 vs. 30 units). The 30 units in a hidden layer also met the minimum number of units required for the unilateral model to support (though not perfectly) the spoken production task (see Supplementary S1 for details).

Fig 2A shows both the developmental learning trajectory of the model before lesion (the intact model) and an example of the recovery profile of the model with a left or right moderate lesion. During the developmental learning period, the model learned high frequency words more accurately and quickly compared to low frequency words. Generalisation to nonwords was very good though lower than performance on words. Then, a moderate lesion was applied to the left or right hidden layer 1 in the model. The representative moderate lesion 50%[0.5] meant that 50% of the units were damaged and noise with the variance of 0.5 was added to the links connecting to and from the left hidden layer 1. After damage, the model was re-exposed to its learning environment for 100,000 presentations to allow for a period of experience-dependent, plasticity-related recovery (based on a re-optimisation of the remaining resources) (44). To mimic a loss of function and missing activation in the damaged brain regions immediately after stroke observed in aphasic patients (46), a period of initial inefficient learning for the surviving units in the damaged layers was

implemented. This meant that for the surviving units in the damaged layers, their learning abilities were initially limited and then gradually regained learning efficiency whereas for the units in the unaffected layers learning efficiency was normal. The initial inefficient learning was implemented by varying unit gain from 0 to 1 in steps of 0.1 over the early stage of retraining (i.e., the first 10,000 presentations for recovery). Note that the model behaved similarly without the implementation of such a period of inefficient learning time (see Supplementary S.2). As can be seen, immediately after left damage, the performance of the model was at floor. Then, the model started to re-organise the computational resources and re-learned the task. In the later stage of recovery, performance gradually increased up to an asymptote (i.e., partial function recovery as found in chronic aphasia). In contrast, the right damage only caused minor disruptions to the performance and it recovered rapidly (i.e., full function recovery akin to transient aphasia).

Obviously, patients may have different lesion severities in the left or right hemisphere, leading to different recovery profiles. To capture this, different levels of damage were applied to the left or right hidden layer 1. Specifically, ten lesion levels were made by damaging hidden units from 10% to 100% with step increment of 10%, plus adding Gaussian noise with variance from 0.1 to 1 with step increment of 0.1 to the links that were connected to and from the target hidden layer. All re-training procedures were the same as described in the previous section. Fig. 2B shows the final recovered performance with different levels of damage to the left or right hidden layer 1. For the left lesion, the recovered performance varied with lesion levels. We divided the models into three lesion groups, 10%[0.1]-30%[0.3] for the mild group, 40%[0.4]-60%[0.6] for the moderate group, and 70%[0.7]-100%[1] for the severe group. The mild group showed the best-recovered performance while the severe group was the worst with the moderate group in the middle. It is worth noting that the relationship between the severity of the left lesion and recovered performance is non-linear, suggesting that the model had developed some resilience to mild damage but, beyond a “tipping point” the effects of damage cannot be overcome through plasticity-related re-learning, leading to

more permanent language impairment as observed in chronic aphasia. For right lesions, the model generally recovered very well regardless of lesion levels. These results demonstrate that, following damage and recovery, performance of the left lesioned model was much more impaired than the right lesioned model. The simulation data were consistent with the patients' studies reported in Gajardo-Vidal et al. (17), showing a stroke in the left hemisphere is more likely to lead profound, chronic language impairment (Figure 2B).

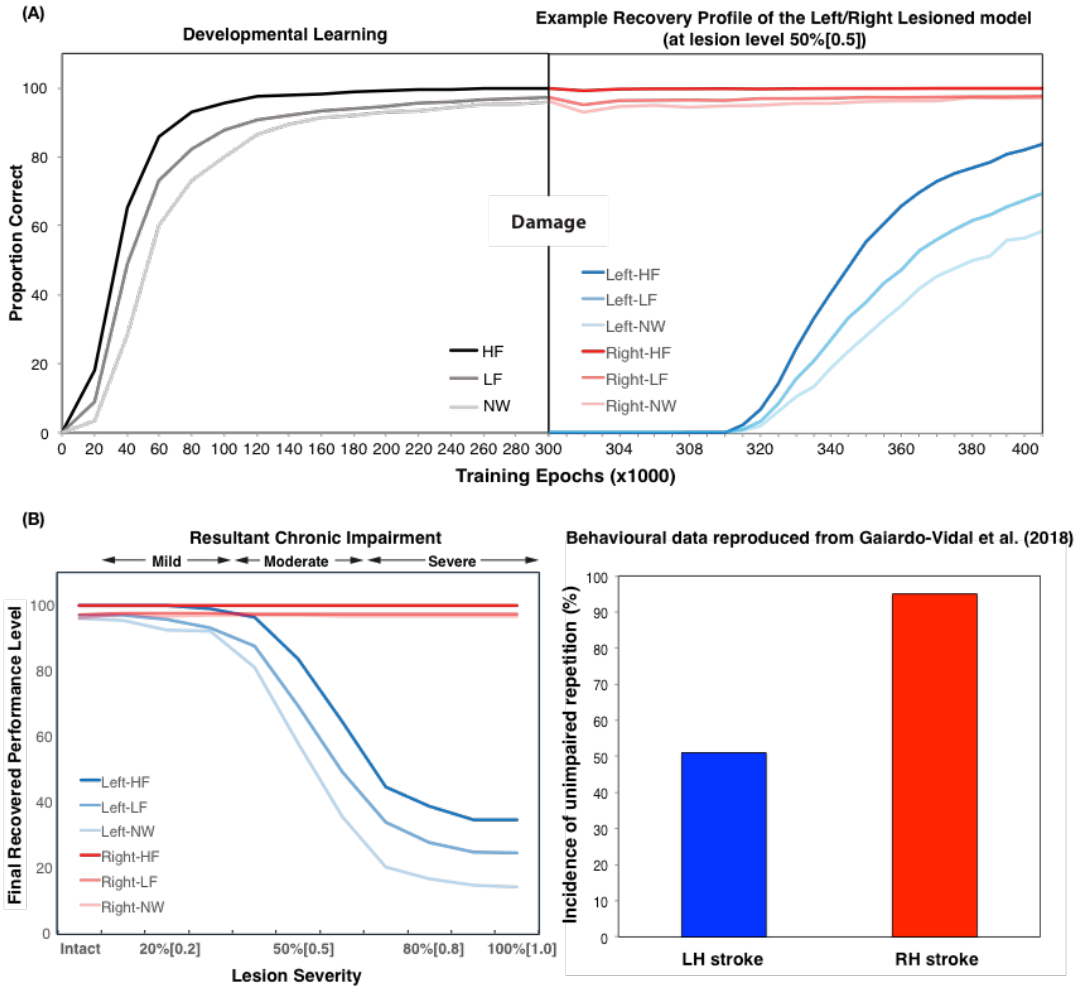

Fig. 2. (A) The developmental learning trajectory of the model before damage, and the recovery profile after damage (Moderate lesion 50%[0.5]) to the left or right hidden layer 1, simulating a left or right hemisphere stroke and recovery. For recovery, the damaged model was re-exposed to its learning environment, in which the model was able to re-optimise the computational resources and gradually relearned the task. Note that the unequally spaced time scales for the re-learning period

were made to clearly demonstrate the model's re-learning in different periods; (B) The recovered performance of the left lesioned model and the right lesioned model as a function of lesion levels (a combination of unit damage and noise – see text for details). 'Intact' means the model without lesion. The behavioural data of unimpaired repetition performance were reproduced from Gajardo-Vidal et al. (17). HF: high frequency words; LF: low frequency words; NW: nonwords; LH: left hemisphere; RH: right hemisphere.

Dynamic patterns of activation shifts in post-stroke aphasia and recovery

An important additional aspect of this study was to investigate the relationship between simulated behavioural performance and underlying metrics of unit function (to mimic functional neuroimaging data). Three levels of the left lesions (20%[0.2], 50%[0.5], and 80%[0.8]) were selected to simulate mild, moderate and severe aphasia. Additionally, the severe right lesion (80%[0.8]) was included to understand what compensated the effects of right damage. Four measures were used to reveal the mechanisms underlying recovery in the damaged model. First, as before, the damaged model's accuracy on word and nonword repetition was used to simulate post-stroke aphasic patients' behavioural performance. Second, we used output unit activation in the left and right pathways as a proxy of the BOLD activation (72) observed in the fMRI studies of post-stroke aphasic patients. Additionally, we investigated whether the model could produce similar activation patterns to that observed in Saur et al. (46) during three phases of language recovery. Specifically, we investigated whether the mild lesioned model could produce: (a) from the acute phase to the subacute phase, an upregulation of the output unit activation in the undamaged left pathway and the right pathway with the highest increase in the right; (b) from the subacute phase to the chronic phase, a downregulation of the output unit activation in the right pathway but the output unit activation in the left remained stable. To simulate that, the re-learning time in the model was divided into three recovery periods (acute, subacute, and chronic) approximating different stages of patient recovery,

and the average output unit activations were computed. Third, we measured the perilesional and contralateral hidden unit activations to examine which undamaged units in the model were upregulated/reformulated to support during recovery. Lastly, we conducted representation similarity analyses (RSA) comparing the activation similarity patterns in the hidden layers to the target output similarity for the words (see the Methods section for details). To our knowledge, there is only one stroke patient study of RSA by Fischer-Baum et al. (68), demonstrating that a patient with a severe lesion to the left VWFA relied more on the right VWFA for orthographic processing, indexed by the RSA similarity scores, while the healthy participants generally relied more on the left VWFA than the right. As this investigation was a single case study of reading (not repetition), we report analogous data based on the closet settings to Fischer-Baum et al. and investigated whether the reliance of the processing shifted from the left to right after severe damage to the left in the model.

In addition to these four measures, two extracted measures were related to the model's relearning, average weight strength and weight change across the hidden layers in the model. Both measures were helpful for understanding how the model re-learned the task during recovery and what the links were between recovery performance and re-learning processes. The data are reported in Supplementary S.3.

Fig. 3 summarises several key phenomena. We can first look at performance accuracy and output unit activation. For the left lesion, the recovered performance of the model aligned with lesion severity with the mild lesion model showing the best performance. Importantly, for the mildest lesion there was a transient pattern of output unit activation shifting from left to right at the early stage of recovery and then back to left at the later stage of recovery, similar to the finding observed in the mild aphasic patients (46). To formally test these observations, we compared the simulation data with Saur et al. (46) by computing the differential output unit activations between the acute and the subacute phases, and between the subacute and the chronic phases. As can be seen in Fig 4A, the output unit activations significantly increased for both the left and right pathways, $t = 8.11$, $p < .001$,

and, $t = 8.56$, $p < .001$, respectively, from the acute to the subacute phase. The increase of activation was numerically higher for the right than for the left pathway. The comparison of the subacute and chronic phases showed a significant decrease of activation in the right pathway, $t = -3.55$, $p = 0.002$, but not in the left pathway ($p = 0.14$). All the statistics were corrected for multiple comparisons.

As for the moderate and severe lesion, the models in Fig 3 showed right lateralised activation patterns, and the recovered performance was worse than that in the mild lesion. In contrast, even after a severe right lesion, accuracy was only slightly disrupted but quickly recovered, and the output activation pattern during recovery largely remained unchanged with a small rise in right output unit activation. Collectively, these simulations mirror the pattern of patient results reported in the literature: good performance is associated with more left lateralised activations while worse performance is associated with more right lateralised activations (5); and, left-right-left changing brain activation patterns are observed in patients with mild brain lesions in the left hemisphere (46).

We further investigated how undamaged perilesional and contralesional units could support recovery. The results showed that, for both mild and moderate lesions, the LH1 perilesional activation initially decreased following damage but then gradually increased during re-learning, reflecting a re-optimisation process. A similar but larger initial decrement followed by a slower increment pattern was observed for LH2 hidden unit activation. For a severe lesion, both the LH1 and LH2 hidden unit activation decreased following damage but did not rise again, presumably because there were insufficient resources available in the LH1 layer for the model to re-optimise. This pattern was also observed for the right severe lesion comparison, where both the RH1 perilesional activation and RH2 hidden unit activation gradually decreased and remained in a low activity level. Turning to contralateral activation, for all severities of left lesion, the contralateral hidden unit activations at RH1 and RH2 were upregulated very quickly following damage. The degree of upregulation was varied and depended on lesion severity, with largest upregulation for the

severe condition. By contrast, for the right severe condition, there was no clear upregulation of the contralateral hidden unit activations at LH1 and LH2.

For the correct interpretation of the relationship between patient behavioural performance and underlying activation, it may be important to note that there were differential associations between model accuracy and the various unit metrics. Figure 3 shows that the RSA measure closely shadowed the changing model accuracy, quite unlike simple unit activation (a proxy to BOLD levels) which show a complex nonlinear relationship. Taking the left moderate lesion as an example, even when the right output unit activation was building up quickly during the initial recovery period, change in model performance was minimal. Subsequently, long after the point when the right output unit activation reached a relatively stable level, there was a much larger and gradual increase in model accuracy. By contrast, the change in the RSA pattern was closely aligned with model performance. Interestingly, although the right output unit activation was higher than the left output unit activation throughout recovery, the RSA results showed the left unit correlation was initially lower than the right unit correlation but returned to a higher level later in recovery. These results suggest that, in the behavioural fMRI studies, BOLD signals and RSA measures may provide different information: although increase unit activations (cf. BOLD increases) are a necessary precursor to behavioural recovery, higher unit activations do not necessarily imply that the units are contributing to improved performance.

To examine, formally, the relationships between model performance with output unit activation and the RSA measure, we conducted correlation analyses. Model performance was correlated with output unit activation and the RSA scores at hidden layers 1 and 2 separately. Correlation analyses were conducted across the developmental learning period in the intact model and the re-learning period in the lesioned model. Results are reported in Table 1. The correlations between output unit activation and model performance were mostly negative in particular for the lesioned conditions, except for the positive correlations for the left output unit activation in the intact

condition and for the right output unit activation in the left severe lesion condition. When considering all intact and lesion conditions, the pattern of change in correlation for output unit activation was difficult to interpret. By contrast, the correlation with the RSA scores were more interpretable; the pattern of correlation change was moderated by lesion severity, revealing the sources of contribution to model performance. For example, left RSA unit correlations were much higher than the right unit correlations in the intact and the left mild lesion condition; conversely, the right unit correlations increased substantially in the left moderate lesion condition and became stronger than the left unit correlations in the most severe left lesion. For the right severe condition, the left unit correlations remained higher than the right unit correlations. In particular, the patterns of the intact model producing a higher RSA unit correlation for the LH1 than the RH1 while the model with a severe left lesion producing a higher RSA unit correlation for the RH1 than the LH1 were conceptually similar to the findings of Fischer-Baum et al. (68), demonstrating the patient with severe damage to left VWFA relied more on the right VWFA for orthographic processing in contrast to the healthy controls relied more on the left VWFA, as shown in Fig. 4B. Obviously, the tasks in the present study and Fischer-Baum et al.'s study are different. Thus, the patterns produced by the model can be considered as predictions for future patient studies in spoken production. Collectively, these results demonstrated that the RSA could potentially provide a more direct measure to relate model performance to the underlying computations. This suggests that, in the studies of post-stroke aphasia, multivariate pattern analyses might be a better way to explore the neural basis of the patients' language behaviour and how this changes during recovery.

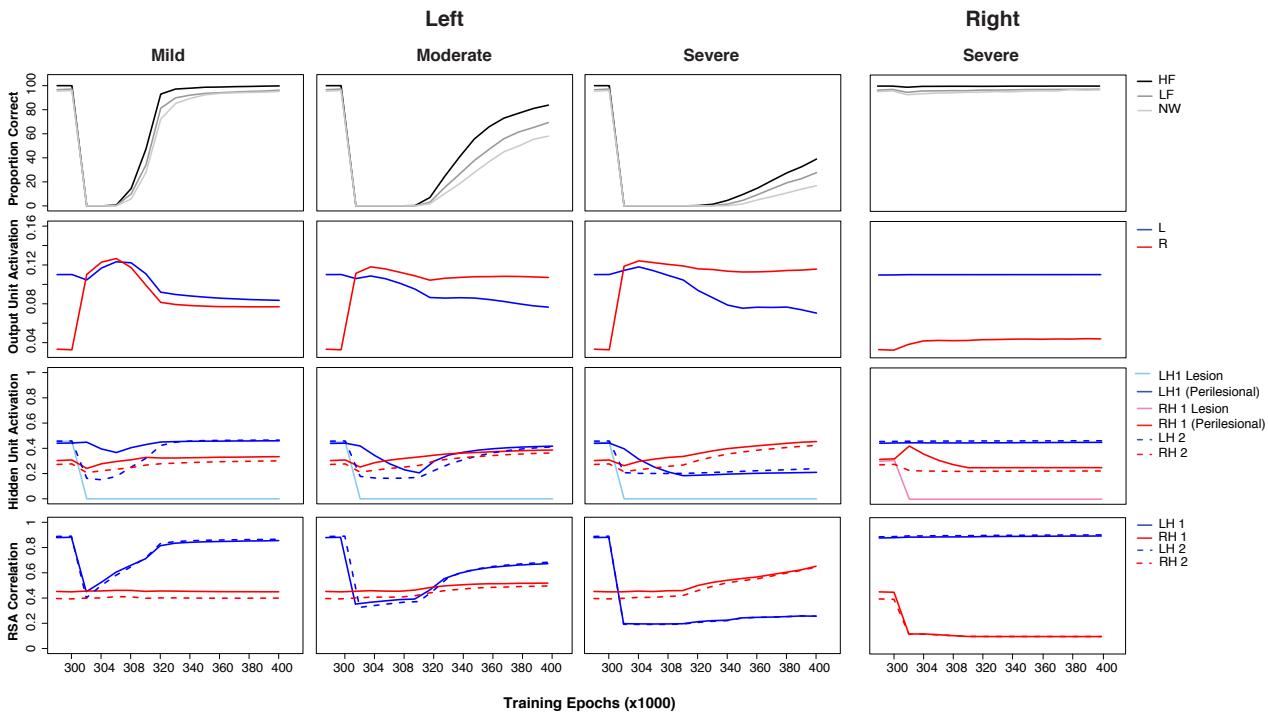

Fig. 3. Simulation patterns of post-stroke aphasia and recovery: left mild (20%[0.2]), left moderate (50%[0.5]), left severe (80%[0.8]) and right severe (80%[0.8]) conditions. The lesion level means the proportion (%) of the units was damaged and the range of noise (bracket) over the links connecting to and from the hidden layer. For each lesion condition, the first panel shows model performance; the second panel shows output unit activation generated from the left and right pathway of the model separately; the third panel shows hidden unit activation for the left and right hidden layers 1 and 2. The activations for lesioned and perilesional units are plotted separately; the last panel shows the RSA scores obtained in the left or right hidden layers 1 and 2 in the model. HF: high frequency words; LF: low frequency words; NW: nonwords; L: left; R: right; LH: left hidden layer; RH: right hidden layer.

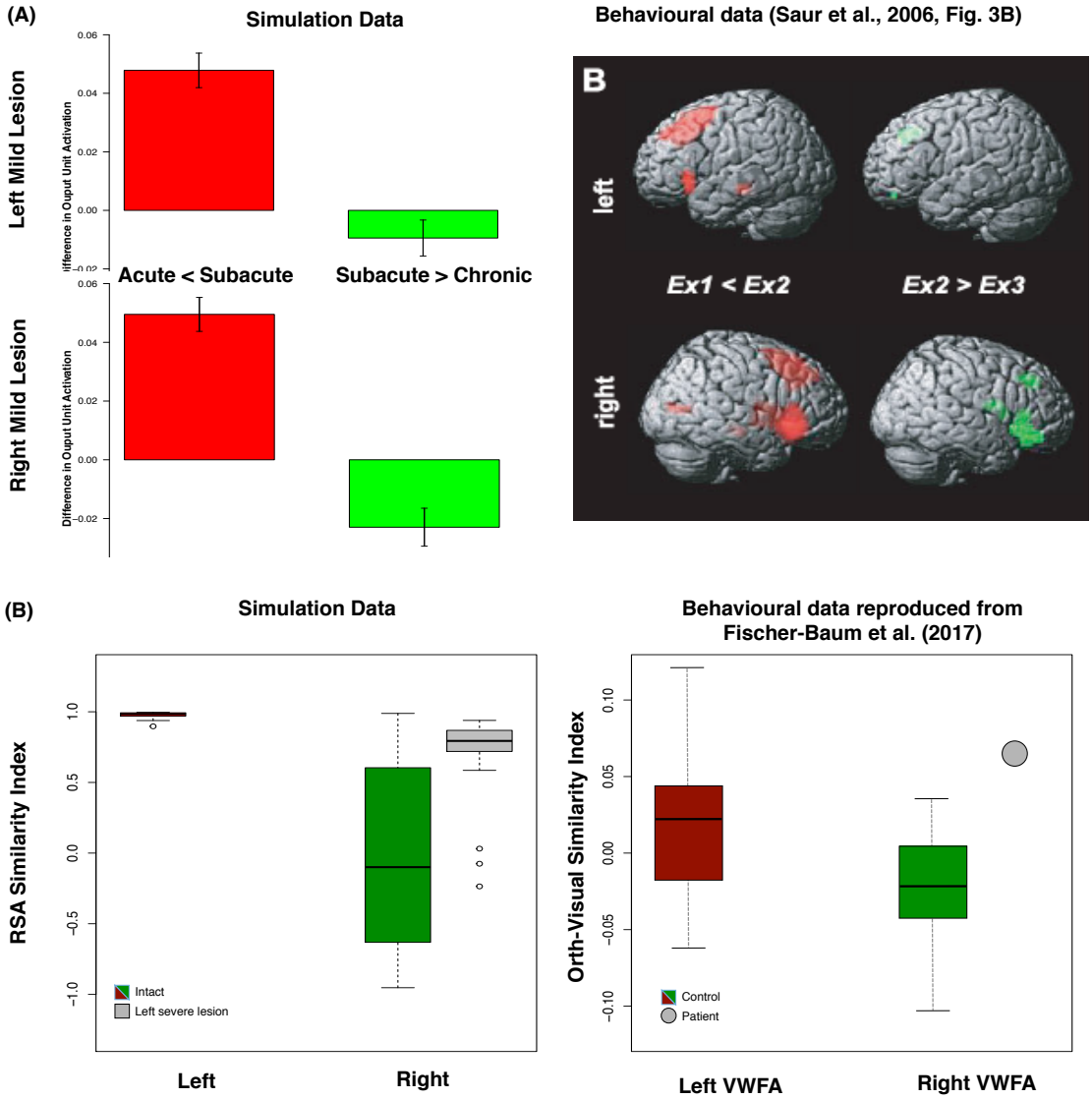

Fig. 4. (A) Upregulation and downregulation in the model with a left mild lesion. Direct comparisons of output unit activations between the acute (300K-301K) and subacute (301K-305K) phases and between the subacute and chronic (305K-340K) phases. The fMRI data of upregulation and downregulation in the language network were from Saur et al. (2006, Fig 3B). Ex1: Acute; Ex2: Subacute; Ex3: Chronic. (B) RSA similarity indices of the left hidden layer 1 for both the intact model and the model with a severe left lesion. The orth-visual similarity indices for both the controls and the patient in the left VWFA and the right VWFA were reproduced from Fischer-Baum et al. (68). VWFA: Visual word form area.

Table 1. The correlations between model performance and output unit activations and RSA scores across the developmental learning period in the intact model and the re-learning period in the lesioned models

	Intact	L Mild	L Moderate	L Severe	R Severe
L Output Unit Act	0.23***	-0.29***	-0.20***	-0.1*	-0.23***
R Output Unit Act	-0.48***	-0.37***	-0.04	0.09*	-0.06
L RSA H2	0.84***	0.92***	0.82***	0.33***	0.42***
R RSA H2	-0.08	-0.04	0.40***	0.69***	-0.08
L RSA H1	0.82***	0.92***	0.81***	0.29***	0.44***
R RSA H1	-0.04	-0.05	0.28***	0.64***	-0.08

$p < .05$; *** $p < .001$; L: left; R: right; Act: activation; RSA: representational similarity analysis.

Interconnectivity between the left and right hemispheres

Thus far, the implemented model did not have interconnections between the left and right pathways. Cortical hemispheres, however, are connected by the corpus callosum as well as various subcortical routes (73). Given that the corpus callosum and interhemisphere connections are complex, a detailed neuroanatomically-constrained simulation is beyond the scope of this study. However, we explored a simplified simulation by adding direct 'homotopic' interconnections between the left and right pathways to investigate whether (a) this changed the patterns of simulated recovery reported above, and (b) if the model would develop transcallosal inhibitory connectivity as proposed in various classical hypotheses (6, 7, 9) (though, to our knowledge, there is no direct evidence of transcallosal inhibitory connectivity outside the motor system). Transcallosal connectivity in the model was implemented as sparse, bidirectional cross-connections between the left and right hemispheres without imposed positive or negative connections. This meant that all the weight connections in the model were allowed to develop freely. As there is no prior knowledge about the density of connectivity between the two hemispheres, we varied two different levels of connectivity sparseness (30% and 70%) of the units in the homologue layers. The training and testing procedures were exactly the same as previously described.

Fig. 5 shows the resulting patterns produced by the left mild, left moderate and left severe and right severe lesioned models with different levels of interconnections. For comparison, the pattern produced by the model without interconnections is also included in Fig. 5. The resulting patterns were very similar to the model with different levels of interconnections. There were transient patterns of output unit activation for the left mild lesion condition but not for more severe left lesion conditions. In addition, the model could recover to a similar accuracy level regardless of the levels of interconnections. But, when the model had more interconnections, it showed a more bilateral pattern following damage and recovery. This suggests that increasing interhemispheric connectivity in the model makes it behave more like a single functional pathway model, with a more even contribution of left and right pathways. This hypothesis was confirmed by the results from the right severe lesion condition, where the model with more interconnections exhibited a more pronounced impairment in the early recovery phase.

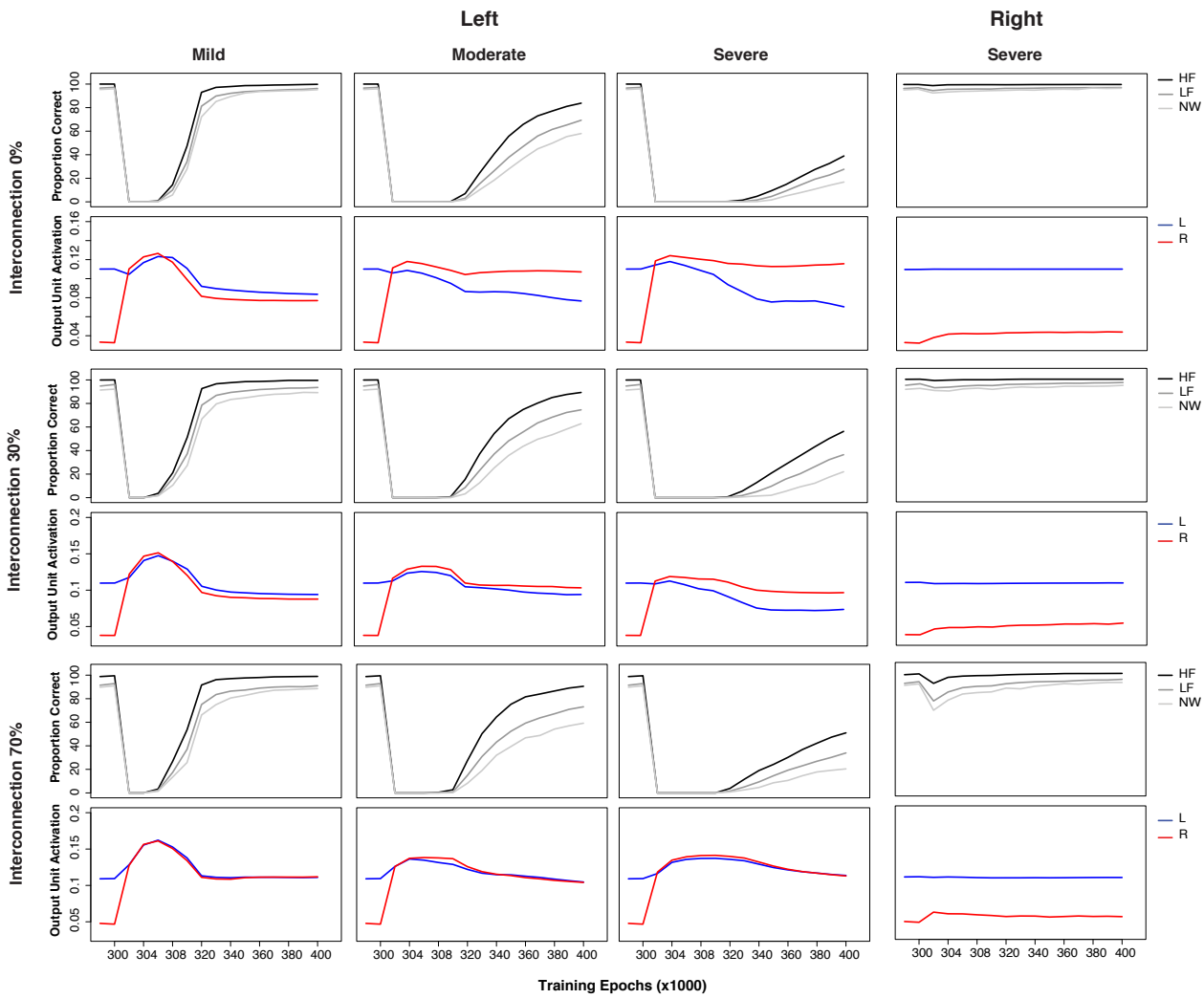

Fig. 5. Simulation patterns of post-stroke aphasia and recovery produced by the model with three levels of interconnections (0%, 30% and 70%) between left and right sides for the left mild (20%[0.2]), left moderate (50%[0.5]), left severe (80%[0.8]) and right severe (80%[0.8]) lesion - conditions. The lesion level means the proportion (%) of the units was damaged and the range of noise (bracket) over the links connecting to and from the hidden layer. For each lesion and interconnection conditions, the first panel shows model performance and the second panel shows output unit activation generated from the left and right pathway of the model separately. HF: high frequency words; LF: low frequency words; NW: nonwords; L: left; R: right.

Discussion

Understanding the brain mechanisms underlying language processing is critical both theoretically and clinically. To tackle various key issues that appear to be contradictory in healthy and impaired language processing, we developed a single, unified neurocomputational model of spoken language production with bilateral pathways. The key features of this modelling work include: the importance of considering healthy and impaired language within an intrinsically bilateral but asymmetric language network; to conceptualise recovery of function after damage as an experience-dependent plasticity-related learning process; and, to provide a platform to assimilate behavioural and neuroimaging data from different populations.

In an otherwise computationally-homogenous language model, an initial imbalance in the processing resources in the left and right hemisphere pathways was sufficient to explain the pattern of data observed in healthy participants and in patients with chronic aphasia. Specifically, the imbalance in processing resources drives an emergent division of labour across the pathways such that the left hemisphere pathway picks up more of the computational work (i.e., each unit, on average, is more highly activated and contributes more to the final spoken output response than each corresponding right hemisphere unit). As a result, the undamaged model shows bilateral but asymmetric "activation" as observed in healthy participants. When this resource imbalance is combined with plasticity-related recovery, the model provides an explanation for why left hemisphere stroke is more likely to result in chronic aphasia than right hemisphere stroke. Plasticity-related recovery reflects a re-optimisation of the remaining connection weights to maximise behavioural performance. This occurs in both 'perilesional' units and the contralateral pathway. The greater computational resources in the left hemisphere means that, when the right hemisphere is damaged, there is greater capacity for the left hemisphere pathway to pick up the extra representational work previously undertaken by the (damaged) right hemisphere pathway (meaning that there is only transient aphasia). The same recovery process occurs following left hemisphere

damage except that (a) the greater left hemisphere resources means that, at least for mild levels of damage, there is still enough spare capacity in the remaining ipsilateral units to pick up the additional computational work (i.e., there is good or recovered function, and left hemisphere activation still dominates, even after mild levels of left hemisphere damage) and (b) there are insufficient resources in the right hemisphere to compensate completely if the left hemisphere damage is too severe. In such circumstances, the model mimics chronic aphasia. In all cases, plasticity-related recovery means that there is a dynamic shift in the division of labour to ipsilateral 'perilesional' and contralateral areas, as is observed in fMRI studies of recovered patients. The model also demonstrates that there can be complex, nonlinear relationships between behavioural performance and levels of unit activation (a proxy for BOLD) whereas the relationship is much more direct when comparing performance to the accuracy of the representations coded in the pathway (implying that MVPA-type imaging analyses may be a better way to assess and track the neural bases of recovery in aphasic patients).

Leftward hemispheric asymmetry has been shown in several brain regions and white matter tracts (35-39). However, there remains some controversy regarding a positive correlation between structural asymmetry and functional lateralisation (29, 40, 41, 74). The discrepancy could be related to individual differences among participants (e.g., age, education, handedness, and gender) or it could be because most studies have relatively small sample sizes (75). In a more controlled computational environment, our bilateral model with differential pathway resources demonstrated a link between hemispheric asymmetry and language lateralisation, consistent with the positive correlation patterns reported in most right-handed healthy subjects (40, 41). The model also shows that this structural difference could be fundamentally important for explaining the patient data. By explicitly incorporating a leftward asymmetric but bilateral structure in the model, the model synthesises the seemingly contradictory patterns observed in both healthy participants and aphasic patients (Fig. 2): specifically, a leftward asymmetric but bilateral patterns in the intact model, and the

much stronger lateralisation picture that is observed in chronic patients after left (aphasic) vs right (recovered) lesions.

Two potential mechanistic frameworks have been proposed for language recovery: degeneracy and variable neuro-displacement (13). Both mechanisms provide the computational bases for the language system to be at least partially resilient to damage and for recovery of function following damage. Recovery can be accomplished by a permanent reformulation of the remaining multiple codes (degeneracy) or upregulation of systems/pathways (variable neuro-displacement), or both. The present neurocomputational model provides a platform to test the two principles. The simulations demonstrate that both mechanisms are not mutually exclusive and they can be utilised as a part of the recovery process. Immediately after dominant pathway damage, the model rapidly upregulates contralesional activation and also starts to re-formulate the perilesional unit contributions. If the perilesional units are capable of re-supporting the function, then later in recovery, both perilesional and contralateral activations are up upregulated; otherwise, the perilesional activation is downregulated and the contralateral activation continues to be upregulated. As such, it would appear from the model that the recovery process follows the two proposed principles but the actual mechanisms involved depend on the level of task engagement by the units before damage and whether there are sufficient resources in the remaining perilesional or contralateral areas to support recovery. As a result, there are differential output activation recovery profiles depending solely on lesion severity. With a mild left lesion, the perilesional units are largely persevered and can be re-formulated for recovery, leading to good recovery and left lateralised output activation patterns. The simulated recovery patterns captured the upregulation and down-regulation of the language network observed in the mild aphasic patients during different stages of recovery (46). With a more severe left lesion to the model, perilesional support is reduced and partial recovery relies mainly on the contralateral units. Accordingly, there is a co-occurrence of slow and imperfect recovered performance with right-lateralised activation patterns. The finding emphasises

the importance of considering lesion severity when interpreting the observations of the association between good recovery and left lateralised brain activation patterns (5, 46) and the association between imperfect recovery and right-lateralised brain activation patterns (76).

The present bilateral model also provides a potential explanation for why the right hemisphere provides some but not perfect language support. The classical right hemisphere hypothesis (RHH) proposes that the right hemisphere is normally suppressed, via transcallosal inhibition, by the dominant left hemisphere system, but it can be released to provide some function after significant left hemisphere damage (6, 7, 9). As noted previously (2, 5, 8, 10-13, 32, 46, 52), the RHH leaves many puzzling questions open, including: how the RH can develop language representations under lifelong suppression; how left and right language systems might contribute to normal function; what bilateral yet asymmetric BOLD activations in healthy participants represents; why this biologically expensive organisation for all people is an optimal solution for the minority of people who happen to suffer from the right kind of brain damage to induce aphasia. The current simulations provide a much more straightforward proposal for the data. The initially bilateral albeit asymmetric system supports healthy function but can partially re-optimise following damage. This can all be achieved without any recourse to notions of juvenile RH language systems and interhemispheric inhibition. Instead, the RH subsystem is less efficient because it has less computational resources and, in turn, learning in the left hemisphere over-shadows that in the right, resulting in the left hemisphere units taking up more of the representational work (Figure 2C). These results follow even without interhemispheric connection but, even if included (Fig 5), then (a) they do not become inhibitory and (b) with increasing connectivity the model evolves into a single functional system. Of course, it should be acknowledged that the connections within corpus callosum are much more complex than the simple parallel connections implemented in the present model. Whilst interhemispheric connections have been shown to be inhibitory within the motor network (58-60), to our knowledge, there is currently no evidence of transcallosal inhibitory in the

language other higher cognitive networks. One study (9), applied TMS to left inferior frontal gyrus in healthy participants during a verbal fluency task, and showed decreased brain activity in the left but increased activity in the right homologue. These findings were interpreted as supportive evidence for transcallosal inhibition from the left to right hemispheres, however, the changes in the effective connectivity between the left and right inferior frontal gyri after TMS was not examined. Alternatively, the upregulation of homologue language areas after brain stimulation could be considered as a form of adaptive plasticity based on an interhemispheric compensatory mechanism (18, 19, 22, 77). For example, a recent study of semantic processing, combining theta-burst stimulation (cTBS) and dynamic causal modelling (DCM) (19) found increased right ventral anterior temporal lobe (vATL) in response to cTBS to the left vATL. The DCM results revealed an increase in the facilitatory drive from the right to the left vATL. There was no evidence of negative inter-ATL connectivity with or without stimulation. Similar results have been reported in another brain stimulation study targeting Broca's area during speech processing (22).

Lastly, the model also investigated multiple measures within a single framework and their sometimes complex relationships. In the model, the performance improvement required both unit activation and fine-tuning weight connections. Immediately after damage, the activation level of the units in the model is generally low. Thus, the first step toward re-learning is to increase the activation level via a generalised weight connection increase. This is then followed by re-optimising weight connections in order to minimise the errors between the target and actual patterns at the output layer. The implication is that fMRI BOLD signals in patients during recovery have an ambiguous interpretation; they could reflect the neural basis for recovered performance or alternatively generalised but untuned activation. The model suggests that multivariate pattern analysis might provide a more direct measure to link recovered performance with neuronal pattern information in different phases of aphasia recovery. This result is consistent with a growing interest in using different types of imaging analyses to investigate the right hemisphere activation patterns in post-

stroke aphasia and how it is related to recovered performance (67, 68). As there has yet to have studies of spoken language production in aphasic patients using RSA, the current simulation data could be considered as model prediction for future developments of the behavioural studies. By extension, the same techniques might also be helpful in clarifying the (dis)advantages of using brain stimulation techniques (TMS or tDCS) to alter brain activation for effective treatments.

To finish we note explicitly that there were at least two deliberate simplifications adopted in the model which can be addressed in future work. First, the present bilateral model focused on speech production along the dorsal pathway. Obviously, there are multiple pathways in the language network (16, 39, 78-81). For example, we have not considered the ventral pathway that includes a semantic system for comprehension, nor does the model specify each layer in the model in corresponding to brain regions involved in language processing. A previous neurocomputational model of language processing (43) demonstrated that a dual-pathway neural network model could simulate different types of aphasia (including receptive and expressive language) based on damage to a corresponding lesion site. Secondly, the psycholinguistic and cognitive science literatures contain many sophisticated models of individual language tasks (typically in relation to healthy performance) that embrace large word corpora and linguistic detail. We deliberately adopted a much more psycholinguistically-simple model so that we could more easily explore the potential relationship between behavioural performance and brain structures/pathways, and thus allow us to distil some principal mechanisms around the emergent division of labour between the two hemispheres in the brain and how this can change in recovery. Future models can merge and elaborate these approaches to provide further systematic investigations, thereby elucidating the neural bases of healthy language and partial recovery in post-stroke aphasia.

Methods

Model architecture

The bilateral model of spoken language production was implemented as a simple recurrent network. It had two pathways to simulate the processing in the left and right hemispheres. Each processing pathway consisted of two hidden layers and one Elman layer for intermediation between input and output phonological layers. The architecture of the model is shown in Fig. 1A. The input phonological layer was connected to the first left and right hidden layers with Elman connections and then to the second left and right hidden layers and then to the single, final output layer. The Elman layer functioned as a memory buffer to temporarily hold activation patterns generated from the previous time ticks (82).

Representation

One hundred three-letter high frequency and one hundred three-letter low frequency monosyllabic words with consonant-vowel-consonant (CVC) structures were included in the training set. Each word was represented by three phoneme slots, with each slot consisting of 25 phonological features (including, for instance, voice, nasal, labial, palatal, round, etc.) following the coding system used in previous modelling work (83, 84). The nonword list comprised three subsets, with 25 items for each of the subsets creating by changing the first consonant, the vowel or the final consonant in a word respectively.

Training and testing

The model was trained on a word repetition task, learning the mapping from phonological to phonological representations. In the first three time ticks, each phoneme was presented in the input layer sequentially. There was no output target until all the phonemes were presented. From the fourth

time tick to the sixth time tick, the model was required to produce all of the phonemes sequentially. Which word was presented to the model was determined by its logarithmic frequency (85). The model was trained with a learning rate of 0.01 using a standard back-propagation algorithm with a negative bias of -2. The weight decay was set to 0.000001. Weight connections in the model were updated after each presentation on the basis of the cross-entropy error computed between the target and the actual activation of the output units. Note that a simple recurrent network generally has a sequential update procedure, which means layers in the network are updated in order. To prevent the order of update from biasing the model's reliance on one pathway, a counterbalance update sequence was used during training.

After 300,000 presentations, the training was halted and the model was tested on the word and nonword repetition tasks. The phonological representation of each phoneme was presented sequentially for the first three time ticks. From the fourth time tick, the activation of units at the output phonological layer was recorded. Error score was measured by the sum of the squared differences between the input representation and its target activation. The accuracy of the model's phonological production was determined by whether the model's actual production was the same as its target phoneme. Twenty versions of the model were trained to prevent the results from generating from a particular set of random initial weights.

Lateralisation index

In the fMRI or PET studies, laterality index can be computed by subtracting the Blood Oxygenation Level Dependent (BOLD) or cerebral blood flow (CBF) signal obtained in the right language areas from the corresponding left language areas and then dividing the score by the sum of the BOLD or CBF signals (41, 70). Thus the higher the score means activation patterns are more left lateralised. For the simulation, we used two measures, functional correlation (44, 71) and output unit activation (72), as a proxy of the BOLD signals. For functional correlation, we recorded the

activation patterns contributed uniquely from the left pathway by lesioning the links between input and the right hidden layer 1; by contrast, the activation patterns contributed uniquely from the right pathway was obtained by lesioning the links between input and the left hidden layer 1. Functional correlation was obtained by correlating the unique activation patterns from each pathway with activation patterns when both pathways were utilised. Regarding output unit activation, it was computed by averaging unit activations contributed uniquely from the left pathway and the right pathway separately. Both were used to replace the BOLD signals in the formula to compute the lateralisation index.

Representational similarity analysis

To conduct representational similarity analyses (86), we first computed a target representational dissimilarity matrix (RDM) based on the correlation distance of target patterns between all of the word pairs. We then computed model representational dissimilarity matrices based on the correlation distance of hidden unit activation patterns between all of the word pairs at hidden layers 1 and 2 in the left and right pathways of the model independently. The RSA correlation scores between the target RDM and the RDMs of hidden unit activations were reported.

Acknowledgements

We are especially grateful to all the patients, families, carers and community support groups for their continued, enthusiastic support of our research programme. This research was supported by an ERC Advanced Grant to MALR (GAP: 670428 - BRAIN2MIND_NEUROCOMP) and Medical Research Council intramural funding (MC_UU_00005/18).

Code availability

The computer code used to run and the simulation and analyse the results will be available on the MRC Cognition and Brain Science Units Data Repository (https://www.mrc-cbu.cam.ac.uk/bibliography/opendata/)

Data availability

The datasets generated and analysed in this study will be available on the MRC Cognition and Brain Science Units Data Repository (https://www.mrc-cbu.cam.ac.uk/bibliography/opendata/)

References

1. Engelter ST, *et al.* (2006) Epidemiology of aphasia attributable to first Ischemic Stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke* 37(6):1379-1384.
2. Barlow T (1877) On a case of double hemiplegia, with cerebral symmetrical lesions. *British Medical Journal* 2(865):103-104.
3. Broca P (1865) Sur le siège de la facultlé du language articulé. *Bull Soc Anthropol* 6:337-393.
4. Wernicke C (Breslau: Cohn and Weigert, 1874) Der aphasische Symptomencomplex, eine psychologische Studie auf anatomischer Basis.
5. Heiss WD, Kessler J, Thiel A, Ghaemi M, & Karbe H (1999) Differential capacity of left and right hemispheric areas for compensation of poststroke aphasia. *Annals of Neurology* 45(4):430-438.
6. Heiss WD & Thiel A (2006) A proposed regional hierarchy in recovery of post-stroke aphasia. *Brain and Language* 98(1):118-123.
7. Karbe H, Herholz K, Halber M, & Heiss WD (1998) Collateral inhibition of transcallosal activity facilitates functional brain asymmetry. *Journal of Cerebral Blood Flow & Metabolism* 18(10):1157-1161.
8. Rosen HJ, *et al.* (2000) Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. *Neurology* 55(12):1883-1894.
9. Thiel A, *et al.* (2006) Direct demonstration of transcallosal disinhibition in language networks. *Journal of Cerebral Blood Flow & Metabolism* 26(9):1122-1127.
10. Cardebat D, *et al.* (2003) Behavioral and neurofunctional changes over time in healthy and aphasic subjects: a PET language activation study. *Stroke* 34(12):2900-2906.
11. Crinion J & Price CJ (2005) Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain* 128(Pt 12):2858-2871.
12. Skipper-Kallal LM, Lacey EH, Xing S, & Turkeltaub PE (2017) Functional activation independently contributes to naming ability and relates to lesion site in post-stroke aphasia. *Human Brain Mapping* 38(4):2051-2066.
13. Stefaniak JD, Halai AD, & Lambon Ralph MA (2020) The neural and neurocomputational bases of recovery from post-stroke aphasia. *Nature Reviews Neurology* 16(1):43-55.

14. Eggert GH (1977) *Wernicke's works on aphasia. A sourcebook and review* (The Hague:Mouton).
15. Geschwind N (1971) Aphasia. *New England Journal of Medicine* 284(12):654-656.
16. Lichtheim L (1885) On aphasia. *Brain* 7:433-484.
17. Gajardo-Vidal A*, et al.* (2018) How right hemisphere damage after stroke can impair speech comprehension. *Brain* 141(12):3389-3404.
18. Binney RJ & Lambon Ralph MA (2015) Using a combination of fMRI and anterior temporal lobe rTMS to measure intrinsic and induced activation changes across the semantic cognition network. *Neuropsychologia* 76:170-181.
19. Jung J & Lambon Ralph MA (2016) Mapping the dynamic network interactions underpinning cognition: A cTBS-fMRI study of the flexible adaptive neural system for semantics. *Cerebral Cortex* 26(8):3580-3590.
20. Lambon Ralph MA, Pobric G, & Jefferies E (2008) Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex* 19(4):832-838.
21. Pobric G, Lambon Ralph MA, & Jefferies E (2009) The role of the anterior temporal lobes in the comprehension of concrete and abstract words: rTMS evidence. *Cortex* 45(9):1104-1110.
22. Hartwigsen G*, et al.* (2013) Perturbation of the left inferior frontal gyrus triggers adaptive plasticity in the right homologous area during speech production. *Proceedings of the National Academy of Sciences* 110(41):16402.
23. Lambon Ralph MA, Cipolotti L, Manes F, & Patterson K (2010) Taking both sides: Do unilateral anterior temporal lobe lesions disrupt semantic memory? *Brain* 133(11):3243-3255.
24. Lambon Ralph MA, Ehsan S, Baker GA, & Rogers TT (2012) Semantic memory is impaired in patients with unilateral anterior temporal lobe resection for temporal lobe epilepsy. *Brain* 135(1):242-258.
25. Lambon Ralph MA, McClelland JL, Patterson K, Galton CJ, & Hodges JR (2001) No right to speak? The relationship between object naming and semantic impairment: Neuropsychological evidence and a computational model. *Journal of Cognitive Neuroscience* 13(3):341-356.
26. Binder JR*, et al.* (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cerebral cortex (New York, N.Y. : 1991)* 10(5):512-528.
27. Bozic M, Tyler LK, Ives DT, Randall B, & Marslen-Wilson WD (2010) Bihemispheric foundations for human speech comprehension. *Proceedings of the National Academy of Sciences* 107(40):17439.
28. Mion M*, et al.* (2010) What the left and right anterior fusiform gyri tell us about semantic memory. *Brain* 133(11):3256-3268.
29. Poeppel D (2014) The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology* 28:142-149.
30. Turkeltaub PE & Coslett HB (2010) Localization of sublexical speech perception components. *Brain and Language* 114(1):1-15.
31. Cogan GB*, et al.* (2014) Sensory–motor transformations for speech occur bilaterally. *Nature* 507(7490):94-98.
32. Blank SC, Scott SK, Murphy K, Warburton E, & Wise RJS (2002) Speech production: Wernicke, Broca and beyond. *Brain* 125(8):1829-1838.
33. Mazoyer B*, et al.* (2014) Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PLOS ONE* 9(6):e101165.

34. Zacà D*, et al.* (2018) Whole-brain network connectivity underlying the human speech articulation as emerged integrating direct electric stimulation, resting state fMRI and tractography. *Frontiers in Human Neuroscience* 12(405).
35. Catani M*, et al.* (2007) Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences* 104(43):17163-17168.
36. Catani M & Mesulam M (2008) The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex* 44(8):953-961.
37. Dorsaint-Pierre R*, et al.* (2006) Asymmetries of the planum temporale and Heschl's gyrus: relationship to language lateralization. *Brain* 129(Pt 5):1164-1176.
38. Geschwind N & Levitsky W (1968) Human brain: left-right asymmetries in temporal speech region. *Science (New York, N.Y.)* 161(3837):186-187.
39. Parker GJ*, et al.* (2005) Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage* 24(3):656-666.
40. Ocklenburg S, Hugdahl K, & Westerhausen R (2013) Structural white matter asymmetries in relation to functional asymmetries during speech perception and production. *NeuroImage* 83:1088-1097.
41. Vernooij MW*, et al.* (2007) Fiber density asymmetry of the arcuate fasciculus in relation to functional hemispheric language lateralization in both right- and left-handed healthy subjects: A combined fMRI and DTI study. *Neuroimage* 35(3):1064-1076.
42. Bain JS, Yeatman JD, Schurr R, Rokem A, & Mezer AA (2019) Evaluating arcuate fasciculus laterality measurements across dataset and tractography pipelines. *Human Brain Mapping* 40(13):3695-3711.
43. Ueno T, Saito S, Rogers TT, & Lambon Ralph MA (2011) Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* 72(2):385-396.
44. Welbourne SR, Woollams AM, Crisp J, & Lambon Ralph MA (2011) The role of plasticity-related functional reorganization in the explanation of central dyslexias. *Cognitive neuropsychology* 28(2):65-108.
45. Postman-Caucheteux WA*, et al.* (2010) Single-trial fMRI shows contralesional activity linked to overt naming errors in chronic aphasic patients. *Journal of Cognitive Neuroscience* 22(6):1299-1318.
46. Saur D*, et al.* (2006) Dynamics of language reorganization after stroke. *Brain* 129(Pt 6):1371-1384.
47. Szaflarski JP, Allendorfer JB, Banks C, Vannest J, & Holland SK (2013) Recovered vs. not-recovered from post-stroke aphasia: The contributions from the dominant and non-dominant hemispheres. *Restorative Neurology and Neuroscience* 31(4):347-360.
48. Szaflarski JP*, et al.* (2011) Poststroke aphasia recovery assessed with functional magnetic resonance imaging and a picture identification task. *Journal of Stroke and Cerebrovascular Diseases* 20(4):336-345.
49. van Oers CA*, et al.* (2010) Contribution of the left and right inferior frontal gyrus in recovery from aphasia. A functional MRI study in stroke patients with preserved hemodynamic responsiveness. *Neuroimage* 49(1):885-893.
50. Warburton E, Price CJ, Swinburn K, & Wise RJS (1999) Mechanisms of recovery from aphasia: Evidence from positron emission tomography studies. *Journal of Neurology, Neurosurgery & Psychiatry* 66(2):155-161.
51. Blank SC, Bird H, Turkheimer F, & Wise RJ (2003) Speech production after stroke: the role of the right pars opercularis. *Annals of Neurology* 54(3):310-320.
52. Basso A, Gardelli M, Grassi MP, & Mariotti M (1989) The role of the right hemisphere in recovery from aphasia. Two case studies. *Cortex* 25(4):555-566.

53. Turkeltaub PE (2015) Brain stimulation and the role of the right hemisphere in aphasia recovery. *Current Neurology and Neuroscience Reports* 15(11):72.
54. Martin PI*, et al.* (2004) Transcranial magnetic stimulation as a complementary treatment for aphasia. *Seminars in Speech and Language* 25(2):181-191.
55. Naeser MA*, et al.* (2005) Improved picture naming in chronic aphasia after TMS to part of right Broca's area: an open-protocol study. *Brain and Language* 93(1):95-105.
56. Turkeltaub PE*, et al.* (2012) The right hemisphere is not unitary in its role in aphasia recovery. *Cortex* 48(9):1179-1186.
57. Winhuisen L*, et al.* (2005) Role of the contralateral inferior frontal gyrus in recovery of language function in poststroke aphasia: A combined repetitive transcranial magnetic stimulation and positron emission tomography study. *Stroke* 36(8):1759-1763.
58. Calautti C, Leroy F, Guincestre JY, & Baron JC (2001) Dynamics of motor network overactivation after striatocapsular stroke: a longitudinal PET study using a fixed-performance paradigm. *Stroke* 32(11):2534-2542.
59. Marshall RS*, et al.* (2000) Evolution of cortical activation during recovery from corticospinal tract infarction. *Stroke* 31(3):656-661.
60. Ward NS, Brown MM, Thompson AJ, & Frackowiak RS (2003) Neural correlates of motor recovery after stroke: a longitudinal fMRI study. *Brain* 126(Pt 11):2476-2496.
61. Bates E*, et al.* (2003) Voxel-based lesion–symptom mapping. *Nature Neuroscience* 6(5):448-450.
62. Butler RA, Lambon Ralph MA, & Woollams AM (2014) Capturing multidimensionality in stroke aphasia: Mapping principal behavioural components to neural structures. *Brain* 137(12):3248-3266.
63. Tyler LK, Marslen-Wilson W, & Stamatakis EA (2005) Dissociating neuro-cognitive component processes: Voxel-based correlational methodology. *Neuropsychologia* 43(5):771-778.
64. Price CJ & Friston KJ (2002) Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences* 6(10):416-421.
65. Blasi V*, et al.* (2002) Word retrieval learning modulates right frontal cortex in patients with left frontal damage. *Neuron* 36(1):159-170.
66. Nardo D, Holland R, Leff AP, Price CJ, & Crinion JT (2017) Less is more: Neural mechanisms underlying anomia treatment in chronic aphasic patients. *Brain* 140(11):3039-3054.
67. Saur D*, et al.* (2010) Early functional magnetic resonance imaging activations predict language outcome after stroke. *Brain* 133(4):1252-1264.
68. Fischer-Baum S, Jang A, & Kajander D (2017) The cognitive neuroplasticity of reading recovery following chronic stroke: A representational similarity analysis approach. *Neural Plasticity* 2017:16.
69. Dilkina K, McClelland JL, & Plaut DC (2008) A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive Neuropsychology* 25(2):136-164.
70. Belin P*, et al.* (1996) Recovery from nonfluent aphasia after melodic intonation therapy: A PET study. *Neurology* 47(6):1504-1511.
71. Chang Y-N, Welbourne S, & Lee C-Y (2016) Exploring orthographic neighborhood size effects in a computational model of Chinese character naming. *Cognitive Psychology* 91:1-23.
72. Chen L, Lambon Ralph MA, & Rogers TT (2017) A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour* 1:0039.
73. Gazzaniga MS (2000) Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain* 123(7):1293-1326.

74. Powell HWR*, et al.* (2006) Hemispheric asymmetries in language-related pathways: A combined functional MRI and tractography study. *NeuroImage* 32(1):388-399.
75. Bishop DVM (2013) Cerebral asymmetry and language development: Cause, correlate, or consequence? *Science (New York, N.Y.)* 340(6138):1230531-1230531.
76. Anglade C, Thiel A, & Ansaldo AI (2014) The complementary role of the cerebral hemispheres in recovery from aphasia after stroke: A critical review of literature. *Brain Inj* 28(2):138-145.
77. Andoh J & Paus T (2011) Combining functional neuroimaging with off-line brain stimulation: modulation of task-related activity in language Areas. *Journal of Cognitive Neuroscience* 23(2):349-361.
78. Hickok G & Poeppel D (2000) Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 4(4):131-138.
79. Hickok G & Poeppel D (2007) The cortical organization of speech processing. *Nature Reviews Neuroscience* 8:393.
80. Rauschecker JP & Scott SK (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12:718.
81. Saur D*, et al.* (2008) Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences* 105(46):18035-18040.
82. Elman JL (1990) Finding structure in time. *Cognitive Science* 14(2):179-211.
83. Chang Y-N, Monaghan P, & Welbourne S (2019) A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language* 108:104025.
84. Harm MW & Seidenberg MS (2004) Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111(3):662-720.
85. Marcus M, Santorini B, & Marcinkiewicz MA (1993) Building a large annotated corpus of English: The Penn Treebank.
86. Kriegeskorte N, Mur M, & Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* 2:4-4.

Supplementary

S1. Explorations of the selection of number of hidden units in the model

To determine the minimum number of units that were required for the model to perform the repetition task, we have developed a unilateral model with different numbers of hidden units. The selection principle followed our assumption that the key difference between the left and right pathways in the model should be quantitative, in terms of differential capacity, rather than qualitative, in terms of function. Thus we ensured that the unilateral model was capable of performing the word and nonword repetition tasks to a satisfactory level (i.e., at least 80% accuracy for both words and nonwords). The architecture of the model and the performance are illustrated in Fig. S1A.

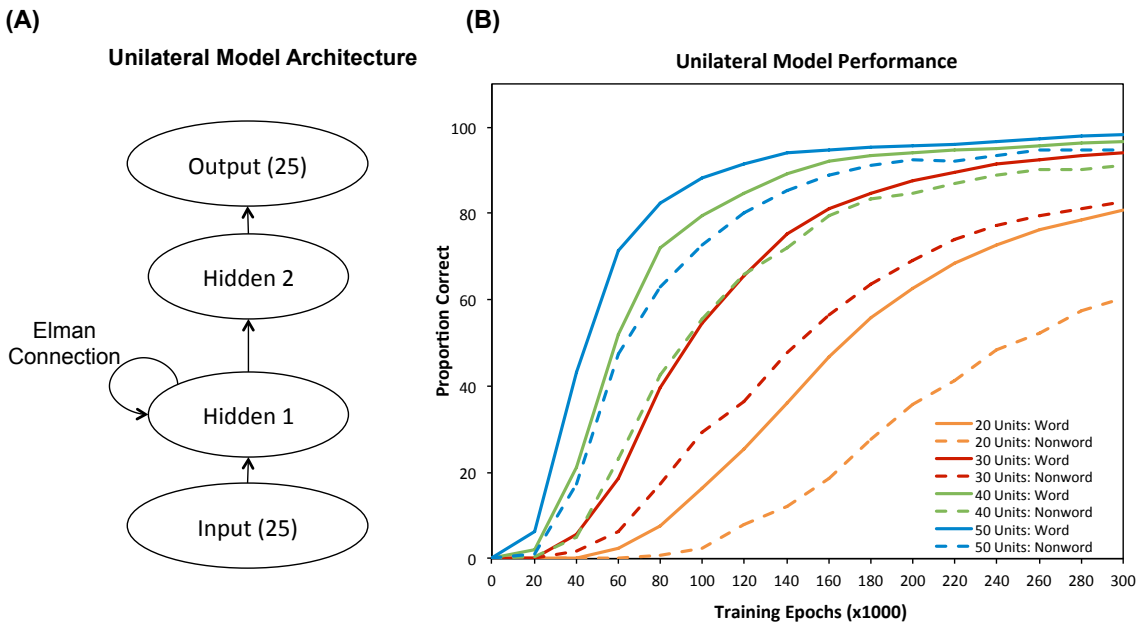

Fig. S1. (A) The architecture of the unilateral model; (B) The performance of the model with different number of units in hidden layers 1 and 2.

We varied the number of hidden layers 1 and 2 concurrently: 20, 30, 40, and 50. The model was trained in the same way as described in the Methods section. After 300,000 presentations, the

model was tested on both the word and nonword repetition tasks. Fig. S1B shows that the model with more hidden units can perform and generalise better. The result demonstrated that the best number of units in each hidden layer was 30 in which the model achieved about 94.1% and 82.8% accuracy on word and nonword repetition tasks respectively. This number was used for the right processing pathway in the left lateralised model reported in the main text.

S2. Explorations of the model's recovery without the implementation of inefficient learning of the surviving units after damage

To simulate behavioural patterns in post-stroke aphasia and recovery, we have trained a damaged model with initial inefficient learning. This was to mimic a loss of function and activation in the damaged brain regions immediately after stroke observed in most patients (46). However, to demonstrate this implementation is not a critical determinant factor to explain different behavioural recovery patterns, we re-trained the damaged model without such an inefficiency period. It means that the surviving units in the hidden layer 1 immediately after damage can learn as efficiently as other units do in the unaffected layers. The levels of damage and the training time of recovery were the same as those described in the Post-Stroke Aphasia and Recovery section. Fig. S2 shows the recovery patterns of the damage model without initial inefficient learning in different lesion conditions.

The resulting performance and output activation patterns were broadly similar to those produced by the model with initial inefficient learning (Fig. 3). When the left lesion was mild, the activation patterns tended to return to be left lateralised during recovery. By contrast, when the left lesion was more severe the activation patterns became right lateralised, and this shift in activation led to relatively poor performance in particular for nonwords. One difference was that the transient pattern from left to right and then back to left previously observed in the left mild lesion condition was less pronounced. However, it was clear that that activity in the right pathway rapidly increased

immediately after damage with decreased activity in the left pathway, though there was no crossover. Regarding all of the other measures, the patterns were very similar to those reported in Fig. 3. These results demonstrate that the simulation without initial inefficient learning could capture the general patterns of recovery in different recovery phases. However, to better characterise the shift in activation patterns in the acute phase, the implementation of initial inefficient learning is critical in simulating a loss of function and activation in the damaged brain regions immediately after stroke observed in most patients (46).

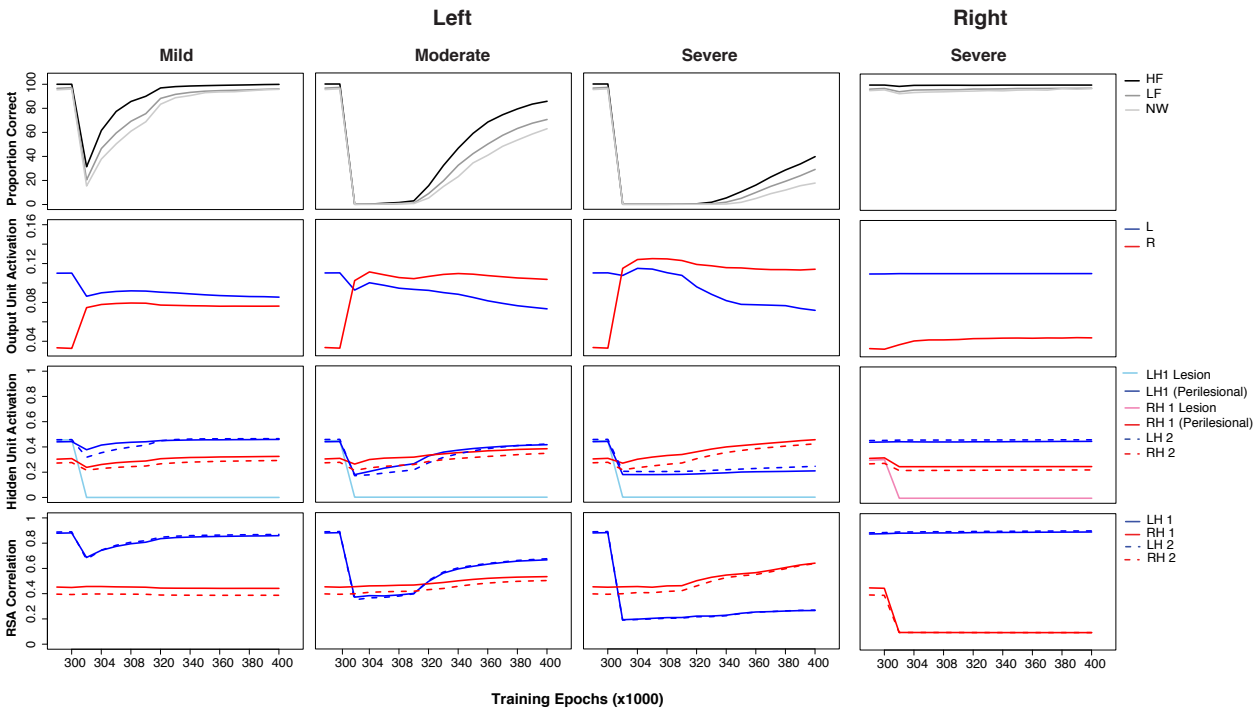

Fig. S2. Simulation patterns of post-stroke aphasia and recovery without initial inefficient training for the surviving units after damage: left mild (20%[0.2]), left moderate (50%[0.5]), left severe (80%[0.8]) and right severe (80%[0.8]) conditions. The lesion level means the proportion (%) of the units was damaged and the range of noise (bracket) over the links connecting to and from the hidden layer. For each lesion condition, the first panel shows model performance; the second panel shows output unit activation generated from the left and right pathway of the model separately; the third panel shows hidden unit activation for the left and right hidden layers 1 and 2. The activations for lesioned and perilesional units are plotted separately; the last panel shows the RSA scores obtained

in the left or right hidden layers 1 and 2 in the model. HF: high frequency words; LF: low frequency words; NW: nonwords; L: left; R: right; LH: left hidden layer; RH: right hidden layer.

S3. A figure for the full simulation patterns of post-stroke aphasia and recover including six measures: model performance, output unit activation, hidden unit activation, weight strength, rate of weight change, and RSA correlation

Six different measures illustrated in Figure S3 were used to reveal the underlying recovery mechanism of the damaged model. In particular, average weight strength and weight change across the hidden layers in the model were useful for us to understand how the model re-learned the task during recovery and what was the link between recovery performance and re-learning processes. For instance, in the left severe lesion condition, the right output unit activation increased rather quickly after damage, and this was also reflected in an initial rise in the rate of weight change. However, the performance had not started to improve at the time. When output unit activation reached steady status, the weights were continues to be updated and the performance was gradually improved. This may indicate two critical steps for re-learning: activation and fine-tuning weight connections. Immediately after damage, the activation level of units in the model is generally low. Thus the first step toward re-learning is to increase the activation level and weight connections, and this is followed by re-optimising weight connections in order to re-learn the task by minimising the errors between the target and actual patterns at the output layer.

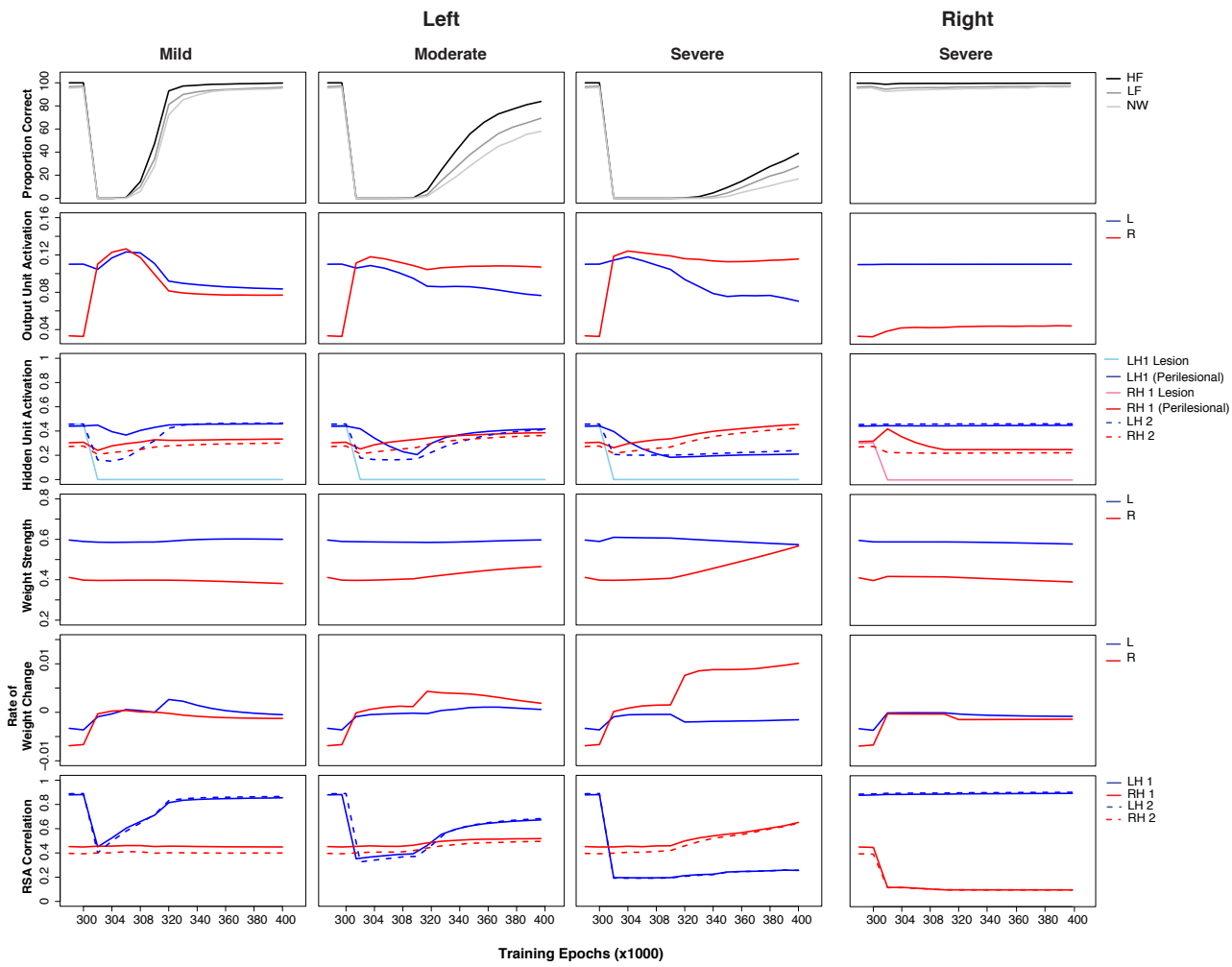

Fig. S3. The full simulation patterns of post-stroke aphasia and recovery: left mild (20%[0.2]), left moderate (50%[0.5]), left severe (80%[0.8]) and right severe (80%[0.8]) conditions. The lesion level means the proportion (%) of the units was damaged and the range of noise (bracket) over the links connecting to and from the hidden layer. For each lesion condition, the first panel shows model performance; the second panel shows output unit activation generated from the left and right pathway of the model separately; the third panel shows hidden unit activation for the left and right hidden layers 1 and 2. The activations for lesioned and perilesional units are plotted separately; the fourth panel shows average weight strength, averaged across all of the connections either in the left or right side of the model; the fifth panel shows the rate of weight change in strength; the last panel shows the RSA scores obtained in the left or right hidden layers 1 and 2 in the model. HF: high

frequency words; LF: low frequency words; NW: nonwords; L: left; R: right; LH: left hidden layer; RH: right hidden layer.