A neurocognitive computational account of word production, comprehension, and repetition in primary progressive aphasia

Ardi Roelofs

A neurocognitive computational account of word production, comprehension, and repetition in primary progressive aphasia

1. Introduction

Aphasia concerns language impairment as a result of brain damage, which is often caused by stroke and much less often by neurodegenerative disease (e.g., Kemmerer, 2015). Poststroke aphasias have been intensively investigated over the past 150 years. Evidence on impaired word production, comprehension, and repetition has informed the verbally described seminal models of Wernicke (1874) and Lichtheim (1885) as well as the model of Geschwind (1970) a century later. Primary progressive aphasia (PPA) is a consequence of neurodegenerative disease, and has been intensively investigated since Mesulam (1982). In a landmark article, Pick (1892/1977) wondered how progressive diffuse atrophy of the brain may lead to focal disorders that resemble those described by Wernicke and Lichtheim. To explain focal progressive disorders of language, he hypothesized that “simple progressive brain atrophy can lead to symptoms of local disturbance through local accentuation of the diffuse process” (p. 40). Modern empirical studies (e.g., Mandelli et al., 2016; Zhou et al., 2012) support this seminal suggestion.

Although poststroke and progressive aphasias show similarities, there are also differences (Grossman & Irwin, 2018; Mesulam, 2007). For example, in terms of language performance characteristics, there does not exist an exact stroke counterpart of semantic dementia (i.e., impaired naming, word comprehension, and conceptual knowledge, alongside spared repetition). Mesulam (2007) stated that “PPA offers a unique experiment of nature for exploring the molecular fingerprints that make the language network a primary disease target and for probing the cognitive architecture of human language as it undergoes a slow but relentless dissolution” (p. 11). However, sophisticated measurements of behavioral performance and brain atrophy have not been matched by advanced computational modeling. While some computationally implemented models, such as the model of Dell et al. (1997, 2013), Lichtheim 2 (Ueno et al., 2011), and WEAVER++/ARC (Roelofs, 2014), have elucidated single-word deficits in poststroke aphasias and in one variant of PPA (viz., semantic dementia), a unified computational account of performance in all major PPA variants is lacking. Computational modeling is a critical tool in testing whether verbal theories are plausible, with the requirement to precisely define the nature of representation and processing, and, in this case, also the nature of impairment (e.g., Mirman & Britt, 2013; Roelofs, 2015; Shallice & Cooper, 2011). Here, a computational account of PPA is presented, based on the WEAVER++/ARC model, which integrates behavioral psycholinguistic, functional neuroimaging, tractography, and aphasiology evidence (WEAVER++/ARC is an acronym standing for Word Encoding by...
Activation and VERification / Arcuate Repetition and Conversation). The account was applied to the three major variants of PPA: nonfluent/agrammatic, semantic (also known as semantic dementia), and logopenic (Gorno-Tempini et al., 2011). The model assumes that PPA arises from a progressive loss of activation capacity in portions of the language network with neurocognitive epicenters that are specific to each PPA variant. Computer simulations were conducted to examine whether the model captures the patterns of relatively impaired and spared word production, comprehension, and repetition performance from representative studies (i.e., Janssen et al., 2022; Savage et al., 2013), at both group and individual patient levels. Also, the model was applied to seminal case series studies (i.e., Gorno-Tempini et al., 2008; Savage et al., 2014) and findings from cluster analyses (Leyton et al., 2015). Moreover, it was examined whether the model captures the patterns of performance with progression of the disease (i.e., Brambati et al., 2015; Mandelli et al., 2016; Rohrer et al., 2013). WEAVER++/ARC and Lichtheim 2 make different predictions regarding naming and comprehension performance in semantic dementia, which were evaluated using empirical evidence.

1.1. Characteristics of primary progressive aphasia and its three variants

Although the three PPA variants present with different foci of neurodegeneration (among widespread cortical thinning) and types of language impairment, disturbances of single-word production, comprehension, or repetition are common (e.g., Gorno-Tempini et al., 2008, 2011; Grossman, 2010, 2018). The consensus criteria for the three PPA variants are described by Gorno-Tempini et al. (2011), and their characteristics are discussed by Kemmerer (2015), among others. The nonfluent/agrammatic variant of PPA is characterized by slow, labored, and halting speech with sound errors (apraxia of speech; over time, patients typically become mute, e.g., Grossman, 2010) and/or agrammatism, atrophy centered on the left inferior frontal gyrus (IFG) and associated white matter, and tau-positive pathology; the semantic variant is characterized by naming and comprehension problems, loss of conceptual knowledge, atrophy centered on the anterior temporal lobe (ATL) bilaterally and associated white matter, and TDP-43-positive pathology; and the logopenic variant is characterized by impaired word retrieval in spontaneous speech and naming, preserved or moderately impaired word repetition and severely impaired phrase and sentence repetition, atrophy centered on the left posterior temporal and inferior parietal cortex and associated white matter, and Alzheimer's disease pathology. In all variants, neurodegeneration affects both gray and white matter (i.e., both cell bodies and fibers of neurons), including long, short, and U-shaped fiber connections (e.g., Brambati et al., 2015). For example, Mandelli et al. (2016) reported that atrophy in the nonfluent/agrammatic variant of PPA progresses in (a) the gray matter of the frontal lobe, with an epicenter in left IFG (i.e., pars opercularis), (b) long-range fiber connections from and to the frontal lobes, such as the arcuate fasciculus that connects posterior superior temporal gyrus (STG) and middle temporal gyrus (MTG) to the IFG, (c) short intralobar frontal connections, such as the frontal aslant tract that connects IFG with the supplementary motor area (SMA), and (d) U-shaped fibers that connect nearby regions within the frontal lobe, such as precentral gyrus and IFG. The behavioral profiles of the three PPA variants concerning single-word production, comprehension, and repetition have been assessed using the Sydney Language Battery (SYDBAT) for English (Savage et al., 2013) and for Dutch using a translated version of this battery (Janssen, 2020; Janssen et al., 2022). Patients with PPA and education- and age-matched healthy controls were compared on picture naming, auditory word comprehension, and word repetition tasks. The target stimuli were thirty imageable nouns of three or more syllables (e.g., butterfly), which were repeated across the tasks. In the picture naming task, participants spoke the name of the object shown in a color photograph (e.g., a butterfly), one at a time. In the word comprehension task, participants saw an array of seven photographs and selected the picture (e.g., butterfly) that best matched the word spoken by the examiner (i.e., “butterfly”). In the repetition task, participants repeated each word spoken by the examiner, one at a time (e.g., repeat “butterfly”). Savage et al. (2013) tested 57 patients with PPA (20 nonfluent/agrammatic, 22 semantic, 15 logopenic) and 54 healthy controls on naming, comprehension, and repetition in English. The control participants performed 89% (27/30, SD = 2.1), 97% (29/30, SD = 1.1), and 100% (30/30, SD = 0.3) correctly on, respectively, the naming, comprehension, and repetition tasks. In the nonfluent/agrammatic variant, picture naming (78% correct; 24/30, SD = 3.3) and repetition (80% correct; 24/30, SD = 6.5) were moderately impaired, while comprehension (94% correct; 28/30, SD = 1.7) was spared. In the semantic variant, naming (23% correct; 7/30, SD = 4.2) was severely impaired and lower than comprehension (63% correct; 19/30, SD = 6.1), while repetition (95% correct; 29/30, SD = 1.5) was preserved (later I provide evidence that the decrease in naming performance is disproportionate). And in the logopenic variant, naming (41% correct; 12/30, SD = 6.8) was rather severely impaired, while comprehension (85% correct; 25/30, SD = 3.0) and repetition (85% correct; 25/30, SD = 5.9) were only moderately affected. The patterns obtained for Dutch patients with PPA by Janssen et al. (2022; Janssen, 2020) are similar to those obtained for English.

A number of studies have examined the progression of atrophy in the brain for the three variants of PPA between a baseline measurement and a follow-up one year later (Brambati et al., 2015; Mandelli et al., 2016; Rohrer et al., 2013). The evidence indicates that the disease progresses over time within variant-specific subnetworks. In the nonfluent/agrammatic variant, atrophy increases in regions including the left IFG and SMA, and integrity decreases in associated fiber tracts, including the frontal aslant tract and the arcuate fasciculus. In the semantic variant, atrophy increases in temporal cortex, and integrity decreases in associated fiber tracts, including the inferior longitudinal fasciculus and uncinate fasciculus. Finally, in the logopenic variant, atrophy increases in temporal cortex, and integrity decreases in associated fiber tracts, including the arcuate fasciculus.

1.2. The WEAVER++/ARC model

The WEAVER++/ARC model integrates behavioral psycholinguistic, functional neuroimaging, tractography, and aphasiology evidence (for a review, see Roelofs & Ferreira, 2019). Computer simulations showed that the model accounts for the outcomes of lesion-deficit analyses that relate damaged brain areas and connections to naming, comprehension, and repetition performance in poststroke aphasia syndromes, including Broca’s, Wernicke’s, conduction, transcortical motor, transcortical sensory, and mixed transcortical aphasia (Roelofs, 2014). Also, the model has been applied to remediation (Roelofs, 2021).

The WEAVER++/ARC model makes a distinction between declarative and procedural aspects of language performance (for a review of the evidence for distinct declarative and procedural memory systems in the brain, see Eichenbaum, 2012). An associative network realizing declarative knowledge is thought to be represented in temporal and inferior frontal areas of the human brain (including Wernicke’s and Broca’s areas). A system of condition-action rules realizing procedural knowledge is assumed to be represented in, among others, the basal ganglia, thalamus, and frontal cortex (including Broca’s area). The associative network is accessed by spreading activation while condition-action rules select among the activated lexical nodes satisfying the task demands specified in working memory (i.e., to name a picture, comprehend a word, repeat a word).

Fig. 1 illustrates the lexical network. It consists of concepts (e.g., CAT [X]), thought to be represented in the ATL bilaterally, lemmas (e.g., cat) in the middle (and probably posterior) part of the left MTG; lexical output forms or morphemes (e.g., cat) in left posterior STG and MTG (Wernicke’s area); output phonemes (e.g., /k/ /a/ and /t/) in left posterior IFG (i.e., Broca’s area); and syllable motor programs (e.g.,
assumes cascading and interactive activation, see Roelofs, 2008, 2014).

In picture naming, activation spreads from lexical concepts to lemmas, between network levels (i.e., different from Levelt et al., 1999, the model specifies that the word is a noun (N; for languages such as Dutch, lemmas also specify grammatical gender). Lemmas also allow for the specification of morphosyntactic parameters, such as number (singular, plural) for nouns, so that the appropriate lexical output form may be retrieved (e.g., singular (cat)).

The condition-action rules mediate top-down influences in conceptually driven word production by selectively enhancing the activation of target lexical concept nodes in the network in order to achieve quick and accurate retrieval and encoding operations. As illustrated in Fig. 1, one source of top-down control is left IFG (see Roelofs, 2014, 2018, for discussion), which is part of frontoparietal and basal ganglia thalamocortical networks underlying domain-general executive control processes and general intelligence (e.g., Duncan, 2010; Posner, 2012). The top-down influences from IFG to the lexical network are mediated by ventral fiber tracts (see Fig. 1), including the uncinate fasciculus and inferior fronto-occipital fasciculus (e.g., Janssen et al., 2020). After external stimulation, activation spreads continuously within and between network levels (i.e., different from Levelt et al., 1999, the model assumes cascading and interactive activation, see Roelofs, 2008, 2014). In picture naming, activation spreads from lexical concepts to lemmas, lexical output forms, output phonemes, and motor programs; in word comprehension, time-varying activation of input phoneme nodes spreads to lexical input forms, lemmas, and lexical concepts; and in repetition, time-varying activation of input phonemes spreads to output phonemes, both directly and via lexical levels (i.e., form or lemma), and to motor programs. Both naming and repetition involve a phonological encoding process that syllabifies the output phonemes and assigns a stress pattern across syllables in polysyllabic words (e.g., butterfly), achieved by condition-action rules. The resulting phonological word representation is used to select the corresponding syllable motor programs (cf. Roelofs, 1997).

1.3. Assumptions about PPA

The model assumes that PPA is the result of progressive degeneration of portions of the lexical network with neuroanatomical loci specific to each PPA variant. As indicated, this assumption of locally intensified degeneration of the brain was first proposed by Pick (1892/1977). Mandelli et al. (2016) argued that the “transneuronal spread of disease occurs selectively within the network anchored to the syndrome epicentre” (p. 2789). As discussed by Dell and colleagues (Dell et al., 1997; Foygel & Dell, 2000; Martin & Dell, 2019), brain damage may reduce the capacity of the lexical network to transmit activation or diminish its capacity to maintain activation over time.

A loss of activation transmission may be functionally implemented as a reduction of connection weights and a loss of activation maintenance as an increased decay rate. In the nonfluent/agrammatic variant of PPA, the neuroanatomical epicenter comprises left IFG (i.e., pars opercularis) and associated white matter, which should functionally affect the subnetwork centered on output phonemes; in the semantic variant, the epicenter comprises bilateral ATL and associated white matter, which should affect the conceptual network in the model; and in the logopenic variant, the epicenter comprises left posterior temporal and inferior parietal cortex and associated white matter, including the arcuate fasciculus running underneath inferior parietal cortex (e.g., Galantucci et al., 2011), which should affect the network centered on lexical output forms and the connections between input and output phonemes in the model. The model assumes that the resulting functional impairment concerns the transmission of activation across network connections (i.e., implemented as a weight decrease for all connections to, within, and from the epicenter) or the maintenance of activation (i.e., an increase in the decay of the activation of nodes in the epicenter). The assumption of a loss in activation transmission has previously been made by other researchers to account for effects of normal aging on language performance (Burke et al., 1991) and the impaired production and repetition performance in poststroke aphasia (Dell et al., 2013). Martin and Dell (2019) observed that patterns of naming errors by individuals with stroke-induced aphasia were better explained by assuming a transmission deficit in some patients and by a maintenance deficit in others.

2. Methods

The network structure and parameter values of the WEAVER+ +/ARC model were the same as in earlier simulations (e.g., Roelofs, 2014).
The simulations were run using a network including words similar to those used by Dell et al. (1997, 2013) and Foygel and Dell (2000), among others. The target was cat and the other words were dog and fish (both semantically related), fog (phonologically related to a semantic alternative, namely dog), and mat (phonologically related to cat). This small network consisted of 5 lexical concept nodes, 5 lemma nodes, 5 lexical input form nodes, 10 input phoneme nodes, 5 lexical output form nodes, 10 output phoneme nodes, 5 syllable program nodes, and corresponding connections. Using other words or including words that were both semantically and phonologically related, such as rat or calf, did not change the simulation outcomes. To examine the effect of varying the size of the lexicon, the simulations were also run with a larger network that contained all the animal names of the SYDBAT (i.e., butterfly, elephant, caterpillar, dinosaur, rhinoceros, hippopotamus, and orangutan). The larger network consisted of 12 lexical concept nodes, 12 lemma nodes, 12 lexical input form nodes, 22 input phoneme nodes, 12 lexical output form nodes, 22 output phoneme nodes, 28 syllable program nodes, and corresponding connections. The simulations with the larger network yielded outcomes similar to those with the smaller network. This suggests that varying the size of the lexicon does not affect the simulation outcomes.

In naming, comprehension, and repetition, information is retrieved from the network by spreading activation. Activation spreads according to

\[ a(m, t + \Delta t) = a(m, t)(1 - d) + \sum \alpha a(n, t). \]

Here, \( a(m, t) \) is the activation level of node \( m \) at point in time \( t \), \( d \) is a decay rate, and \( \Delta t \) is the duration of a time step in ms. The rightmost term denotes the amount of activation that \( m \) receives between \( t \) and \( t + \Delta t \), where \( a(n, t) \) is the output of neighbor \( n \) (equal to its level of activation). The factor \( r \) indicates the strength of the connection between nodes \( m \) and \( n \) (i.e., its weight). To implement the transmission deficit assumption, atrophy severity was simulated by manipulating connection weights \( r \) at specific network loci (cf. Dell et al., 2013). Atrophy was assumed to decrease the value of \( r \), which corresponds to a loss in activation transmission. To implement the maintenance deficit assumption, atrophy severity was simulated by manipulating decay rate \( d \) at specific network loci. Damage was assumed to increase the value of \( d \), which corresponds to a loss of activation maintenance.

The simulations started by providing external activation to lexical concepts for naming and to input phonemes for repetition and comprehension. Activation was then allowed to spread in \( \Delta t = 25 \) ms steps for 2 sec and the mean activation of nodes was computed. Condition-action rules were assumed to select nodes depending on the task. For example, lexical concept, lemma, lexical output form, output phoneme, and syllable motor program nodes are selected for naming. In displaying results, I concentrate on the ultimate target nodes for each task (naming, comprehension, repetition). The ultimate targets were syllable motor program nodes for naming and to input phonemes for repetition, and lexical concept nodes for comprehension. Selection of targets by condition-action rules may sometimes fail (see San José et al., 2021, for discussion). Errors may occur in the model when the selection condition of an alternative node is incorrectly taken to be satisfied or a rule selects a wrong node among the activated ones. The probability of an error is proportional to the difference in activation between target and closest alternatives. Thus, a reduction of this difference due to damage would correspondingly reduce performance accuracy. For each of several degrees and loci of damage, the difference in mean activation between target and closest alternative in the damaged network was computed and expressed as a percentage of the normal activation difference. With smaller activation differences, selection takes longer and errors are more likely to occur, so lower percentages will correspond to poorer performance. For naming and repetition, the activation difference concerned syllable program nodes, and for comprehension, the difference concerned lexical concept nodes.

It should be emphasized that the model does not actually make errors (nor was it designed to). What is modeled is the difference in activation between the target and closest alternatives, and a reduction is assumed to reduce accuracy. Consequently, an additional mechanism would be required for errors to actually be produced in the model, and hence impaired performance to actually be simulated. Whereas the naming and repetition tasks require the production of spoken words, the comprehension task asks for selection of the best matching picture from arrays of photographs containing the target item and six foils, involving a performance element that is not present for naming or repetition.

To summarize, the independent variables in the simulations were the locus of external input to the network (i.e., concepts for naming and input phonemes for repetition and comprehension) and the locus and severity of damage. The dependent variables were the difference in activation between target and closest alternatives relative to normal at the level of lexical concepts (for comprehension) and syllable program nodes (for naming and repetition). With smaller activation differences, selections take longer and accuracy decreases.

The weight decrease in the simulations concerned the connections to and from the output phonemes in the nonfluent/agrammatic variant; the connections to, within, and from the conceptual network in the semantic variant; and the connections to and from lexical output forms and between input and output phonemes in the logopenic variant. The decay increase concerned the output phonemes in the nonfluent/agrammatic variant; the concepts in the semantic variant; and the lexical output forms in the logopenic variant. The values of the weight decrease and decay increase parameters that provided the best fit between model and data were obtained through a grid search, that is, an exhaustive search through the parameter space, varying between minimal and maximal damage. Weight decrease was varied between 0.99 and 0.0 \((<r>)\) and decay increase between 1.01 and 1.66 \((<d>)\), both in steps of 0.01. The largest value of decay increase (i.e., 1.66) represents full decay, reducing \(a(m, t)\) to zero during a time step. The search aimed to obtain the parameter value for each PPA variant and individual patient that minimizes the mean absolute difference (i.e., mean absolute error, MAE) between simulated and empirical performance for naming, comprehension, and repetition. The MAE is simpler and more directly interpretable than the commonly used root mean-square error (Willmott & Matsuura, 2005), with a higher MAE denoting a worse fit. To indicate goodness of fit between model and data, MAEs and Pearson correlation coefficients are reported.

The simulations with the WEAVER+++/ARC model were computationally implemented using the C programming language and the programming environment of Microsoft Visual C++ 2017. The source code of the simulation program is available from the Open Science Framework at https://osf.io/ue4bn/ or from the author.

3. Results and discussion

3.1. Performance accuracy as a function of atrophy severity

Fig. 2 displays how performance accuracy varies as a function of weight decrease (top panel) and decay increase (bottom panel) in simulations of naming, comprehension, and repetition in the three PPA variants. The figure shows that weight decrease and decay increase tend to have similar effects, but there are also differences. In the nonfluent/agrammatic variant, naming and repetition worsen with weight decrease or decay increase, while comprehension is relatively preserved. In the semantic variant, naming and comprehension get worse with weight decrease or decay increase, with naming being more severely disrupted than comprehension, while repetition is preserved. And in the logopenic variant, weight decrease worsens naming and repetition, with naming being more severely affected than repetition, while comprehension is relatively preserved; with decay increase, naming gets worse, while repetition and comprehension are preserved. These patterns of relatively impaired and spared performance in the different PPA
variants correspond generally to what is empirically observed (i.e., Brambati et al., 2015; Janssen et al., 2022; Mandelli et al., 2016; Rohrer et al., 2013; Savage et al., 2013), although there are also discrepancies between model and data.

Some aspects of the patterns of performance that are observed empirically (Janssen et al., 2022; Savage et al., 2013) seem more similar to the decreased performances that result from weight decreases (top panel of Fig. 2) than from decay increases (bottom panel). For example, in the nonfluent/agrammatic variant, many individuals eventually become mute (e.g., Grossman, 2010), which is consistent with an ultimate accuracy of 0% on naming and repetition tasks. Fig. 2 shows that this may result from a weight decrease but not from a decay increase. In the logopenic variant, neither comprehension nor repetition is likely to remain at ceiling (85% correct for both in Savage et al., 2013) and naming can be very poor (41% correct in Savage et al.), which may be accommodated by a weight decrease but not a decay increase. Therefore, the predictions for the empirical studies depicted and discussed in the next sections were derived in simulations assuming a weight lesion, whereas the results for a decay lesion are only briefly mentioned and reported in more detail in the Supplementary material.

![Fig. 2. Performance accuracy as a function of weight decrease (top panel) and decay increase (bottom panel) in WEAVER++/ARC simulations of naming, comprehension, and repetition in the nonfluent/agrammatic, semantic, and logopenic variants of primary progressive aphasia.](image)

![Fig. 3. Performance accuracy in the nonfluent/agrammatic, semantic, and logopenic variants of primary progressive aphasia for naming, comprehension, and repetition: Empirical group averages for English (Savage et al., 2013) and Dutch (Janssen et al., 2022) and WEAVER++/ARC simulation results. The error bars represent 95% confidence intervals. For each panel, the estimated weight decrease (e.g., 0.91) and mean absolute error (MAE) are displayed. N = number of patients.](image)
3.2. Behavioral profiles of the PPA variants

Fig. 3 shows the WEAVER++/ARC simulation results for naming, comprehension, and repetition in the three PPA variants, along with the empirical group averages for English (Savage et al., 2013) and Dutch (Janssen, 2020; Janssen et al., 2022). According to the model, naming and repetition are moderately affected while comprehension is largely preserved in the nonfluent/agrammatic variant; naming is more severely affected than comprehension while repetition is preserved in the semantic variant; and naming is rather severely disrupted while comprehension and repetition are only moderately impaired in the logopenic variant. This was also observed for the logopenic variant (N = 34) by Ramanan et al. (2020) using the SYDBAT. The MAE measures of fit between simulated and empirical group data are generally small, averaging 4.9% and ranging between 0.8% and 9.2% (see Supplementary Table 1 for details). The correlation between model and data is \( r = 0.95, p < .001 \). Similar outcomes were obtained when assuming a decay lesion (Supplementary Table 1), with an average MAE of 5.5%, range between 0.6% and 15.3%, and correlation of \( r = 0.96, p < .001 \). The poor naming observed by Savage et al. for the logopenic variant cannot be captured well by a decay lesion in the model, yielding an MAE of 15.3%.

Although the MAEs are generally small, there are also places where the model does not do so well in fitting the data, as revealed by the error bars representing 95% confidence intervals for the empirical data (cf. Cumming, 2014). In particular, the model underestimates naming in the semantic variant (English study) and repetition in the logopenic variant (both studies), which is further discussed below.

According to the model, the presence of a conceptual deficit in the semantic variant leads to impaired naming and comprehension (which both depend, to different degrees, on conceptual representations) and spared repetition (which does not require conceptual representations). The presence of a phonological output deficit in the nonfluent/agrammatic variant leads to impaired naming and repetition (which both depend on output phonemes) and spared comprehension (which does not require output phonemes), while the absence of a phonological output deficit in the semantic and logopenic variants leads to relatively spared repetition. However, naming is impaired in these variants for different reasons, which is attributable to the conceptual deficit in the semantic variant and to a lexical output deficit in the logopenic variant. As a consequence of the lexical output deficit in the logopenic variant, naming tends to be more impaired than repetition (i.e., naming requires the lexical output forms, whereas repetition does not).

To assess conceptual knowledge, Savage et al. (2013) and Janssen et al. (2022) used a fourth task of the SYDBAT that involves picture–picture matching: Participants saw a target picture and had to select a closely related picture from a set of four options. As with word comprehension, performance was impaired in the semantic variant of PPA and relatively spared in the nonfluent/agrammatic and logopenic variants. This shows that the loss of conceptual knowledge in the semantic variant is the same across input modalities (cf. Patterson et al., 2007).

Leyton et al. (2015) presented evidence from cluster analyses (N = 21) for functional and neuroanatomical heterogeneity in the logopenic variant. Their analyses revealed three clusters differing in single-word task performance. The clusters did not differ in demographic characteristics or duration of symptoms, except that Cluster 1 scored higher on overall cognitive performance as assessed by the Addenbrooke’s Cognitive Examination-Revised (AICE-R). The other two clusters, which did not differ. Cluster 1 (N = 10) showed the typical logopenic profile of impaired naming and relatively spared comprehension and repetition (as shown in Fig. 3 for group averages), with left temporoparietal atrophy and Alzheimer pathology. Cluster 2 (N = 6) additionally showed impaired word comprehension (i.e., comprehension was worse than repetition, different from what is shown in Figs. 2 and 3), but differed from the semantic variant of PPA in that there was no loss of conceptual knowledge. Moreover, there was extension of atrophy in temporal cortex but no marked atrophy of the ATL. Cluster 3 (N = 5) displayed a severe version of the typical logopenic pattern (as shown in Fig. 3), with much impaired naming and repetition alongside relatively spared comprehension. Atrophy was observed mainly in the left STG, including the planum temporale, and in the insula, which are areas involved in phonological processing.

Simulations revealed that the WEAVER++/ARC model does a reasonable job of capturing the performance profiles of the three clusters, as shown in the top panel of Fig. 4 (see Supplementary Table 2 for details). Clusters 1 and 3 could be captured by assuming weight lesions of different severities, as displayed in Fig. 2 and in line with the different scores on the ACE-R. However, the simulation of Cluster 1 underestimates repetition. To account for Cluster 2, the connections between lemmas and concepts were assumed to be lesioned (i.e., intended to capture the more extended temporal cortex atrophy but with sparing of the ATL and thus conceptual knowledge) rather than the connections between input and output phonemes. The MAE measures of fit between simulated and empirical data are generally small, with an average of 4.0% and range between 2.3% and 6.0%. The correlation between model and data is \( r = 0.98, p < .001 \). Similar outcomes were obtained when assuming a decay lesion, with an average MAE of 4.2%, range between 0.4% and 7.2%, and correlation of \( r = 0.96, p < .001 \).

The simulations discussed so far concerned group averages. However, it is also important to assess whether the model can capture performance of single patients (e.g., Dell et al., 1997). Although many single-case studies of PPA have been reported in the literature (e.g., Mesulam, 1982; Westbury & Bub, 1997), case series studies seem less common. Seminal reports concern the case series studies of Savage et al. (2014) on the semantic variant of PPA (N = 5) and of Gorno-Tempini et al. (2008) on the logopenic variant (N = 6). These studies showed that the patterns of behavioral performance and brain atrophy of the patients within a PPA variant were all similar. Researchers of PPA therefore maintain that it makes sense to examine average behavioral and imaging data at the group level (e.g., Brambati et al., 2015; Galantucci et al., 2011; Mandelli et al., 2016; Ramanan et al., 2020; Rohrer et al., 2013), different from what is often assumed in the stroke literature (e.g., Foygel & Dell, 2000). For an extensive discussion of this issue, I refer to Shallice and Cooper (2011) and Shallice (2015). To examine whether the model captures performance of single cases, it was applied to the individual patients in the studies of Savage et al., Gorno-Tempini et al., and Janssen et al. (2022). This was also the approach to stroke-induced aphasia taken by Dell and colleagues, which enabled detailed testing of their model, including highlighting the patterns that the model could not simulate.

Simulations of Savage et al. (2014) and Gorno-Tempini et al. (2008) revealed that the model captures the findings on the individual patients reasonably well, as shown by the middle and bottom panels of Fig. 4 (see Supplementary Table 3 for details). Note that there are no 95% confidence intervals as each data point reflects a single performance. Assuming a weight lesion, the MAE measures of fit between simulated and empirical data of the single cases are generally small, with an average MAE of 5.7% and range from 1.7% to 7.0%. The correlation between model and data is \( r = 0.96, p < .001 \). Similar outcomes were obtained when assuming a decay lesion, with an average MAE of 7.7%, range between 0.7% and 17.0%, and correlation of \( r = 0.95, p < .001 \). The poorest fits (with an MAE of 17.0%) are obtained for cases of the semantic variant with very low naming performance (3–7% correct), which cannot be accommodated by a decay lesion (see Fig. 2). Although they are generally rare, the model either underestimates naming (cases 1 and 2) or overestimates repetition (cases 3–5) for the semantic variant, and it underestimates (cases 1, 2, and 4) or overestimates (case 5) repetition for the logopenic variant. Note that due to measurement error and patient-specific variability, over- and underpredictions are expected even if the model’s predictions are correct.

The top panel of Fig. 5 shows the patterns of performance on the naming, comprehension, and repetition tasks of the 45 individual patients in the study of Janssen et al. (2022) together with the group
averages for the three PPA variants. Using box plots to determine outliers observations in the patient scores for each task within a PPA variant (e.g., Tukey, 1977), only 3 out of the 135 data points were deemed to be outliers, denoted by numbers (e.g., #10). The figure reveals that for the nonfluent/agrammatic variant, the pattern for the group corresponds to the individual patterns. This also holds for the semantic variant, except for one patient (case 10) with extremely disrupted comprehension and another patient (case 2) with severely disrupted repetition. For the logopenic variant, the pattern for the group also corresponds to the individual patterns, except for one patient (case 2) presenting with exceptionally disrupted repetition.

The bottom panel of Fig. 5 shows the WEAVER++/ARC simulation results for all 45 individual patients in the study of Janssen et al. (2022), assuming a weight lesion. The performance scores for each patient (denoted by dots and numbers with the tasks being color coded) are plotted at the lesion value with the lowest MAE along with the model predictions (for details, see Supplementary Table 4). For the nonfluent/agrammatic variant, the curves for naming and repetition in the model are identical. Overall, the model succeeds reasonably well at simulating the performance patterns of the individual cases. The average MAEs across patients for the nonfluent/agrammatic, semantic, and logopenic variants are, respectively, 4.9%, 3.7%, and 7.3%. The overall correlation between model and individual patient data is $r = 0.93$, $p < .001$. Similar outcomes were obtained when assuming a decay lesion (Supplementary Table 5), with average MAEs across patients for the nonfluent/agrammatic, semantic, and logopenic variants of, respectively, 4.8%, 6.3%, and 5.2%, and a correlation of $r = 0.92$, $p < .001$. Although the MAEs are generally small, it should be noted that the model overestimates naming by at least 10% (and often more) for half of the cases of the nonfluent/agrammatic variant (see Supplementary Table 4). For the logopenic variant, the model underestimates repetition in the majority of cases. Although there are cases of the semantic variant whose performance is not well simulated, such as repetition of case 2 and naming and comprehension of case 4, in general, there is a less clear pattern of model discrepancy than with the other two variants.

The model yields the outlying pattern of case 10 of the semantic variant (i.e., very poor naming and comprehension with preserved repetition) when the weight lesion is severe (MAE = 3.4%). Fig. 2 shows that this pattern cannot occur with a decay lesion (MAE = 29.3%). Although case 2 of the semantic variant presents with exceptionally poor repetition, which is overestimated by the model, it still fits the overall pattern of performance rather well (MAE = 8.4%). The pattern of case 2 of the logopenic variant is obtained with a severe weight lesion in the model (MAE = 10.7%). The poorest fit is obtained for logopenic case 11 (MAE = 17.5%), whose naming was severely disrupted (40% correct) along with preserved comprehension (100% correct) and repetition (93% correct). As Figs. 2 and 5 show, poor naming goes together with lowered repetition performance under the assumption of a weight lesion in the model, which does not hold for this patient. This may suggest that the mapping of input onto output phonemes is intact in this patient, which would yield the observed pattern of performance in the simulation (with an MAE of 2.4%). Alternatively, this patient may have a decay lesion (which yields an MAE of 8.9%), which impairs naming while preserving comprehension and repetition, as shown by Fig. 2.

The goodness of fit between the model and the individual patient data of Janssen et al. (2022) is worse for the logopenic variant (MAE ranging between 1.3% and 17.5% for a weight lesion) than for the nonfluent/agrammatic variant (MAE ranging between 0.8% and 11.2%)
and the semantic variant (MAE ranging between 0.8% and 9.3%), indicating some heterogeneity in performance profiles across the logopenic cases. This is in line with the evidence from the cluster analyses by Leyton et al. (2015) that the logopenic variant presents with functional and neuroanatomical heterogeneity. The patient variability results in a difficulty in the model’s ability to simulate the patterns. For example, the underestimation of repetition in the majority of logopenic patients of Janssen et al. would suggest that the connections between input and output phonemes are not impaired in these patients. However, in other patients the connections seem to be impaired, like in cluster 3 of Leyton et al. and in Brambati et al. (2015), as shown below.

### 3.3. Progression of disease

Fig. 6 shows the WEAVER++/ARC simulation results for baseline and follow-up performance in the three PPA variants, along with the empirical group averages of Brambati et al. (2015), Mandelli et al. (2016), and Rohrer et al. (2013). Whereas Brambati et al. tested all three PPA variants, Mandelli et al. and Rohrer et al. concentrated on single variants with larger sample sizes. Overall, performance becomes worse over time, while preserving the patterns of relatively impaired and spared performance for each PPA variant. In the nonfluent/agrammatic variant, naming and repetition are moderately affected, while comprehension is preserved. In the semantic variant, naming is more severely disrupted than comprehension. However, different from the simulations, repetition performance is also lowered somewhat. This suggests that the progression of atrophy toward posterior temporal cortex in the semantic variant also affects the mapping of input onto output phonemes, which impairs repetition in the model. In the logopenic variant, naming is rather severely disrupted, while comprehension and repetition are only moderately impaired. The best fit between model and data indicates a weight decrease (or decay increase) from baseline to follow-up assessment. The MAE measures of fit between simulated (weight decrease) and empirical data are generally small, with an overall average of 4.9% and range from 0.2% to 8.7% (for details, see Supplementary Tables 6 and 7). The correlation between model and data is $r = 0.94$, $p < .001$. Similar outcomes were obtained when assuming a decay lesion (Supplementary Tables 6 and 7), with an average MAE of 7.0%, range from 0.1% to 17.2%, and correlation of $r = 0.88$, $p < .001$.

### 3.4. Comparison with other models

Performance is typically worse for naming than comprehension, regardless of the PPA variant (see Figs. 3–6). We saw that WEAVER++/ARC captures this finding, but an important question is whether the same holds for other implemented models. The Lichtheim 2 model (Ueno et al., 2011) has been applied to the semantic variant (but not to the other two PPA variants). In the model, meaning (thought to be represented in ventrolateral anterior temporal cortex) is mapped via two layers of hidden nodes (thought to be represented in anterior STG/STS and in IFG opercularis-triangularis) onto articulatory output (thought to be represented in insular-motor cortex) in picture naming, while auditory word input (thought to be represented in primary auditory areas and surroundings) is mapped via two layers of hidden nodes (thought to be represented in middle STG/STS and in anterior STG/STS) onto meaning (as indicated, thought to be represented in ventrolateral anterior temporal cortex) in comprehension. Computer simulations with Lichtheim 2 conducted by Ueno et al. themselves showed how performance accuracy in semantic dementia varies as a function of lesion severity, which concerns the proportion of damaged links in the layer representing meaning (in ventrolateral anterior temporal cortex). These simulations revealed that naming and comprehension are affected to the same extent regardless of severity value, contrary to the empirical findings. Thus, WEAVER++/ARC, but not Lichtheim 2, captures the
finding that naming is much worse than comprehension in the semantic variant of PPA.

In WEAVER++/ARC, the naming pathway through the lexical network is longer than the comprehension pathway. In naming, conceptual representations are mapped via lemmas, lexical output forms, output phonemes, and syllable motor programs onto articulatory output, whereas in comprehension, auditory input is mapped via input phonemes, lexical input forms, and lemmas, onto conceptual representations. Moreover, a conceptual deficit is carried through the entire processing pathway in naming but not in comprehension. As a consequence, naming is more severely disrupted than comprehension in the model. In contrast, the naming and comprehension pathways are equally long in Lichtheim 2. As indicated, in naming, meaning is mapped via two layers of hidden nodes onto articulatory output, whereas in comprehension, auditory input is mapped via two layers of hidden nodes onto meaning. Moreover, Lichtheim 2 uses distributed representations of meaning and sounds, so that a deficit is carried differently through the network than in WEAVER++/ARC. Due to these differences between the two models, damage has a differential impact, impairing naming more than comprehension in WEAVER++/ARC but not in Lichtheim 2.

It may be argued that one should be cautious about saying that naming is more severely impaired than comprehension. These are different tasks, with comprehension having a chance element (i.e., involving selection of a target from a set of seven photographs) that is different from naming, and healthy controls also performing differently across tasks. For example, in the study of Janssen et al. (2022), controls performed the naming and comprehension tasks with 90% and 96% accuracy, respectively. To assess whether naming is disproportionately impaired in the semantic variant, performance ratios of naming relative to comprehension were compared between groups (cf. Janssen et al., 2020). The ratio was 0.33 for the semantic variant and 0.94 for the controls. Statistical analysis revealed that the ratios differed between groups, Welch $\text{t}(12.48) = 9.89$, $p < .001$, indicating that the decrease in naming performance in the semantic variant is disproportionate. WEAVER++/ARC, but not Lichtheim 2, captures this finding.

4. General discussion

The present article reports on an extension of the WEAVER++/ARC model to the three major variants of PPA. Following a seminal suggestion of Pick (1892/1977) and modern empirical insights, PPA is assumed to arise from a progressive loss of activation capacity in portions of the
Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2022.105094.

References

Brambati, S. M., Amici, S., Racine, C. A., Neuhaus, J., Miller, Z., Ogár, J., Drönkers, N., Miller, B. L., Rosen, H., & Gorno-Tempini, M. L. (2015). Longitudinal gray matter matter contraction in three variants of primary progressive aphasia: A tensor-based morphometry study. *NeuroImage. Clinical, 8*, 345–355. https://doi.org/10.1016/j.nicl.2015.01.013

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*(5), 542–579. https://doi.org/10.1016/0749-596X(91)90026-G

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dell, G. S., Schwartz, M. F., Nourari, N., Fasenijat, O., & Coslett, H. B. (2013). Voxel-based lesion–parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition, 128*(3), 380–396. https://doi.org/10.1016/j.cognition.2013.06.007

Duncan, J. (2010). How intelligence happens. *New Haven, CT: Yale University Press.*

Eichenbaum, H. (2012). The cognitive neuroscience of memory: An introduction (2nd Ed.). *Oxford, UK: Oxford University Press.*

Foygel, R., & D. G. S. (2009). *Simulations of impaired lexical access in speech production.* *Journal of Memory and Language, 62*(2), 182–216. https://doi.org/10.1016/j.jmla.2009.07.001

Galantucci, S., Tartaglia, M. C., Wilson, S. M., Henry, M. L., Filippi, M., Agosta, F., Drönkers, N. F., Henry, R. G., Ogár, J. M., Miller, B. L., & Gorno-Tempini, M. L. (2011). White matter damage in primary progressive aphasia: A diffusion tensor tractography study. *Brain, 134*(10), 3011–3029. https://doi.org/10.1093/brain/awr299

Geschwind, N. (1970). The organization of language and the brain. *Science, 170*(3961), 940–944. https://doi.org/10.1126/science.170.3961.940

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain’s language pathways. *Cerebral Cortex, 18*(11), 2471–2482. https://doi.org/10.1093/cercor/bhn101

Gorno-Tempini, M. L., Brambati, S. M., Gines, V., Ogár, J., Drönkers, N. F., Marcone, A., & Miller, B. L. (2008). The logopenic/phonological variant of primary progressive aphasia. *Neurology, 71*(12), 1227–1234. https://doi.org/10.1212/01.wnl.0000320506.79811.da

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogár, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Drönkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., & Gorno-Tempini, M. L. (2011). Classification of primary progressive aphasia and its variants. *Neurology, 76*(11), 1006–1014. https://doi.org/10.1212/200103e318210136e6

Grossman, M. (2010). Primary progressive aphasia: Clicopathological correlations. *Annual Review of Neuroscience, 33*, 377–396. https://doi.org/10.1146/annurev-neuro-070110-134853

Grossman, M., & Irwin, D. J. (2018). Primary progressive aphasia and stroke aphasia. *Continuum, 24*(3), 745–767. https://doi.org/10.1212/CON.0000000000000618

Hagoopt, P., & Indefrey, P. (2014). The neurolinguistics of language beyond single words. *Annual Review of Neuroscience, 37*, 347–362. https://doi.org/10.1146/annurev-neuro-070113-013847

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience, 8*(5), 393–402. https://doi.org/10.1038/nrn2113

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Article 255 Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Janssen, N. (2020). Staying connected as we speak: Behavioral and tractography evidence from health and neurodegenerative disease [Unpublished doctoral dissertation]. *Nijmegen: Radboud University.*

Janssen, N., Roelofs, A., Margans, M., Sierpowska, J., Kessels, R., & Piai, V. (2020). How the speed of word finding depends on ventral tract integrity in primary progressive aphasia. *NeuroImage. Clinical, 29*, Article 102450. https://doi.org/10.1016/j.nicl.2020.102450

Janssen, N., Kessels, R. P. C., Mars, R. B., Llera, A., Beckmann, C. F., & Roelofs, A. (2021). Disociating the functional roles of arcuate fascicular tractubs in speech production. Manuscript submitted for publication.

Janssen, N., Roelofs, A., Van den Broek, E., Holleman, M. A., In de Braek, D. M. J. M., Piguet, O., & Kessels, R. P. C. (2002). The diagnostic value of language screening in primary progressive aphasia: Validation and application of the Sydney Language Battery. *Journal of Speech, Language, and Hearing Research, 45*(1), 200–214. https://doi.org/10.1044/2021_JSLHR-21-00024

Kemmerer, D. (2015). *Cognitive neuroscience of language.* New York, NJ: Psychology Press.
A. Roelofs

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER model of word-form encoding in speech production. *Brain and Language, 222*(1), 1–38. https://doi.org/10.1017/S0278739314001776

Lichtheim, L. (1885). On aphasia. *Brain, 7*(4), 433–484.

Roelofs, A. (2015). Simulation of impairment: Commentary on Nickels, Rapp, and Kohnen (2015). *Cognitive Neuropsychology, 32*(7–8), 427–430. https://doi.org/10.1080/02643294.2016.1154832

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition, 172*, 59–72. https://doi.org/10.1016/j.cognition.2017.12.007

Roelofs, A. (2020). Self-monitoring in speaking: In defense of a comprehension-based account. *Journal of Cognition, 31*(1), Article 18. https://doi.org/10.5334/joc.61

Roelofs, A. (2021). Phonological cueing of word finding in aphasia: Insights from simulations of immediate and treatment effects. *Aphasiology, 35*(2), 169–185. https://doi.org/10.1080/02687038.2019.1686748

Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. In P. Hagoort (Ed.), *Human language: From genes and brains to behavior* (pp. 35–50). Cambridge, MA: MIT Press.

Rohrer, J. D., Caso, F., Mahoney, C., Henry, M. H., Rosen, H. J., Rabinovici, G., Rosser, M. N., Miller, B., Warren, J. D., Fox, N. C., Ridgway, G. R., & Gorno-Tempini, M. L. (2013). Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia. *Brain and Language, 127*(2), 121–126. https://doi.org/10.1016/j.bandl.2012.12.008

San José, A., Roelofs, A., & Meyer, A. S. (2021). Modeling the distributional dynamics of attention and semantic interference in word production. *Cognition, 211*, Article 104636. https://doi.org/10.1016/j.cognition.2021.104636

Savage, S., Hsieh, S., Leslie, F., Foxe, D., Piquet, O., & Hodges, J. R. (2013). Distinguishing subtypes in primary progressive aphasia: Application of the Sydney language battery. *Dementia and Geriatric Cognitive Disorders, 35*(3–4), 208–218. https://doi.org/10.1159/000346389

Savage, S. A., Piquet, O., & Hodges, J. R. (2014). Giving words new life: Generalization of word retraining outcomes in semantic dementia. *Journal of Alzheimer’s Disease, 40*(2), 309–317. https://doi.org/10.3233/JAD-131826

Shalllice, T. (2015). Cognitive neuropsychology and its vicissitudes: The fate of Caramazza’s axioms. *Cognitive Neuropsychology, 32*(7–8), 385–411. https://doi.org/10.1080/02643294.2015.1131077

Shalllice, T., & Cooper, R. P. (2011). The organisation of mind. Oxford, UK: Oxford University Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (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. https://doi.org/10.1016/j.neuron.2011.09.013

Wernicke, C. (1874). *Der aphatische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis [The aphasic symptom complex: A psychological study on anatomical basis]*. Breslau: Cohn und Weigert.

Westbury, C., & Bub, D. (1997). Primary progressive aphasia: A review of 112 cases. *Brain and Language, 60*(3), 381–406. https://doi.org/10.1006/brln.1997.4140

Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language, 122*(3), 190–198. https://doi.org/10.1016/j.bandl.2012.04.005

Zhou, J., Gennatas, E. D., Kramer, J. H., Miller, B. L., & Seeley, W. W. (2012). Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron, 73*(6), 1216–1227. https://doi.org/10.1016/j.neuron.2012.03.004