A graded, multidimensional approach to comparing aphasia across aetiologies

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2020

Ruth Ursula Ingram

School of Biological Sciences
Division of Neuroscience and Experimental Psychology
# Table of contents

List of figures ........................................................................................................... 5
List of tables .............................................................................................................. 5
Abstract ..................................................................................................................... 7
Declaration ............................................................................................................... 8
Copyright Statement ............................................................................................... 8
Acknowledgements ................................................................................................. 9
Chapter 1 - Introduction ........................................................................................ 10
  Thesis aim ............................................................................................................. 10
  The nosology of aphasia ..................................................................................... 10
    Post-stroke aphasia ......................................................................................... 11
    Primary progressive aphasia .......................................................................... 12
  Comparisons of aphasia across aetiologies ....................................................... 14
  Reconceptualising variance along multiple phenotypic spectra ...................... 19
  Specific aims per chapter ............................................................................... 24
Chapter 2 - Graded, multidimensional intragroup and intergroup variations in primary progressive aphasia and post-stroke aphasia ................................................................. 26
  Statement of contribution ................................................................................. 26
  Abstract .............................................................................................................. 27
  Introduction ....................................................................................................... 28
  Methods ............................................................................................................ 31
    Patients .......................................................................................................... 31
    Neuropsychological assessments ................................................................. 32
    Data analysis ................................................................................................. 32
  Results ............................................................................................................. 35
    Demographics ............................................................................................... 35
    Principal component analysis ...................................................................... 35
    Intergroup comparisons ............................................................................... 40
    Intragroup graded variation ......................................................................... 43
  Discussion ....................................................................................................... 47
  Supplement .................................................................................................... 52
Chapter 3 - Assessing the validity of the novel Mini-Linguistic State Examination in post-stroke aphasia

Statement of contribution

Abstract

Introduction

Method

Participants

Test batteries

Data analysis

Results

Demographics

Convergence of subtype classifications

Sensitivity

Convergent validity of MLSE subtests

Discussion

Chapter 4 - Exploring transdiagnostic brain-behaviour relationships in aphasia using the Mini-linguistic State Examination

Statement of contribution

Abstract

Introduction

Methods

Participants

Mini-linguistic state examination

Data analysis

Results

Demographic details

Profiles of impairment on the MLSE

Shared principal components of aphasia

Subgroup atrophy maps

Voxel-based correlational methodology

Discussion

Patterns of linguistic and neural abnormality
Directions for future research................................................................. 159
Conclusion................................................................................................. 162
References.................................................................................................. 163

Word count = 56,476

List of figures

Figure 2.1 – Intergroup comparison of the underlying dimensions of variance in PSA and PPA. ................................................................. 40
Figure 2.2 - Regions of the shared multidimensional space of PSA and PPA occupied by each diagnostic subtype................................................................. 42
Figure 2.3 - Data-driven diagnostic cut-off values for semantic dementia................................................................. 44
Supplementary Figure 2.4 – Principal components extracted for post-stroke aphasia. .................. 52
Supplementary Figure 2.5 – Principal components extracted for primary progressive aphasia................................................................. 54
Supplementary Figure 2.6 – Principal components extracted for the unified principal component analysis of primary progressive aphasia and post-stroke aphasia. .................. 56
Figure 3.1 – Sensitivity of the MLSE to mild deficits................................. 72
Figure 3.2 – Sensitivity of the MLSE to levels of impairment........................ 75
Figure 4.1 – MLSE performance per subgroup........................................... 95
Figure 4.2 – Shared multidimensional space of language impairments across aetiologies. 100
Figure 4.3 – Patterns of neural abnormality relative to controls, across aetiologies. ............. 104
Figure 4.4 – Significant clusters for phonology, semantics, and syntax, extracted using voxel-based correlational methodology. ........................................... 107
Figure 5.1 – Graded variation along principal dimensions for posterior cortical atrophy. .... 125
Figure 5.2 – Graded variation along a spectrum of visual processing impairments in typical AD and posterior cortical atrophy.................................................. 129
Figure 6.1 – Commonalities among principal component analyses in the thesis. .............. 141

List of tables

Table 2.1 – Demographic details per subtype of the post-stroke aphasia and primary progressive aphasia cohorts. .................................................. 35
Table 2.2 – Distribution of misclassifications between clinical and data-driven diagnostic PPA groups. ........................................................................ 46
Supplementary Table 2.3 – Iterative sweep through Unified PCA multidimensional space to find cut-offs giving optimal diagnostic selectivity. ........................................... 57
Supplementary Table 2.4 – Distribution of misclassifications between clinical and data-driven diagnostic PSA groups.......................................................................................................................... 58

Table 3.1 – Subtypes of post-stroke aphasia.............................................................................................................. 62
Table 3.2 – Demographic details of post-stroke aphasics and healthy controls................................................. 69
Table 3.3 – Convergence of subtype classifications derived using the MLSE and BDAE...................... 71
Table 3.4 - Sensitivity of the MLSE to mild deficits. .......................................................................................... 73
Table 3.5 - Convergent validity of the MLSE subtests. ....................................................................................... 77
Table 4.1 – Demographic details of all participants. ......................................................................................... 92
Table 4.2 – Principal component analysis of MLSE performance across aetiologies. .................. 98
Table 4.3 – Anatomical locations of clusters for phonology, semantics and syntax. ............................ 106
Table 5.1 – Demographic details per diagnostic group. ................................................................................. 121
Table 5.2 – Principal component analysis for posterior cortical atrophy. ...................................................... 123
Table 5.3 – Principal component analysis for typical Alzheimer’s disease...................................................... 127
Table 6.1 – Shared principal dimensions across PSA and PPA................................................................. 140
A graded, multidimensional approach to comparing aphasia across aetiologies

Ruth Ursula Ingram
The University of Manchester
For the degree of Doctor of Philosophy, March 2020

Abstract
Aphasia can arise due to stroke (post-stroke aphasia, PSA) or neurodegeneration (primary progressive aphasia, PPA). Two clinical and theoretical issues of relevance to PSA and PPA are addressed in this thesis. First, there have been few detailed direct comparisons across the full phenotypic ranges of PSA and PPA. Second, graded differences between aphasic subtypes and within-subtype heterogeneity suggest that phenotype differences observed across patients might reflect graded variations across multidimensional aphasic spectra rather than mutually exclusive diagnostic categories. Thus, the aim of this thesis was to compare directly the full ranges of PSA and PPA using a graded multidimensional framework to map out graded intragroup and intergroup differences. Chapter 2 uncovered graded variation in PSA and PPA along phonological, semantic, and fluency-related dimensions. Plotting all cases in the transdiagnostic shared multidimensional space revealed that ‘fluent’ PSA were often less fluent than ‘non-fluent’ PPA. Classification analysis showed that semantic dementia occupied a unique region of this multidimensional space due to the selectivity of the semantic deficit. Chapter 2 highlighted the importance of a shared test battery for this transdiagnostic approach. The Mini-Linguistic State Examination (MLSE) is a novel clinical assessment tool which is in development to address the lack of a standardised linguistic assessment tool valid for aphasia across aetiologies. Chapter 3 conducted the first formal validation of the MLSE and compared it to established aphasia tests, finding that the MLSE is highly sensitive to the presence of language impairments in PSA. Chapter 4 compared linguistic impairments across PSA and PPA on the MLSE, extending the multidimensional framework from Chapter 2 by extracting brain-behaviour relationships that underly variance in aphasia across aetiologies. Language impairments across PSA and PPA varied along three key dimensions: phonology, semantics and syntax, which were related to unique neural correlates in: left hemisphere superior and posterior temporal lobe, bilateral anterior temporal lobes, and left hemisphere subcortical regions and periventricular white matter, respectively. Chapter 5 extended the multidimensional framework to typical amnestic Alzheimer’s disease (AD) and atypical visual variant AD. These groups varied gradedly along a spectrum of visual processing impairments; typical AD with mild visual impairments and mild visual variant AD overlapped in the middle of this spectrum. Overall, the transdiagnostic, multidimensional approach applied in this thesis highlights that considering patient variance along continuous dimensions instead of categorical systems has benefits for comparing directly the full phenotypic space of different disorders and relating principal behavioural dimensions to unique neural correlates of structural brain abnormalities.
Declaration
No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement
I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DoculInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Acknowledgements

I would first like to express my gratitude to all the people who participated in this research, in particular the unfailingly kind and generous people who gave their time and energy to undergo neuropsychological testing of just the things that they find most difficult – thank you for your resilience and selflessness. Thanks are next due to the research teams who tested these people and who generously made the extensive neuropsychological datasets available for analysis in this thesis - to Professor Sebastian Crutch and his team at University College London who collected the posterior cortical atrophy dataset, to Dr Seyed Sajjadi and Professor Karalyn Patterson for the primary progressive aphasia dataset from the Cambridge longitudinal study of primary progressive aphasia, and to Professor Peter Garrard, Dr Nikil Patel and Dr Katie Petersen for the Mini-linguistic State Examination dataset. My thanks also go to Dr Anna Woollams and Professor Jason Warren for examining my thesis.

I would also like to express my gratitude to my supervisory team, Dr Gorana Pobric, Professor Matt Lambon Ralph, and Dr Ajay Halai. Gorana, thank you for all the times you helped me ‘cut to the chase’ and helped me learn to be pragmatic. Ajay, thank you for the innumerable times you provided technical support (linear indexing in four-dimensional matrices!) and for sharing your methodological expertise with me. Matt, thank you for teaching me how to frame my research ideas so that I don’t go down too many research-rabbit holes.

A whole host of friends and family have also been invaluable to my PhD journey. Special thanks go to Dr Natalie Busby, Dr Sarah Martin and my little sister Lana, for keeping me positive, curious, and most importantly, sane. Extra special thanks go to my boyfriend Ben for making me laugh, even when I thought I didn’t want to. Lastly, I would like to thank my parents. I am incredibly lucky to have a father who is also doing a PhD as I am writing up this thesis – thank you Dad, for all the proof-reading, stats-searching, for loop-coding, and MS Office hacking that you have helped me with throughout this PhD. I am also extremely fortunate to have a mother who is hands-down the best mind-mapper in the world. Thank you, Mum, for taking care of me whilst I took care of the thesis.
Chapter 1 - Introduction

This thesis is written in the Journal Format, with Chapters 2-5 being empirical manuscripts which have been or could be submitted for publication. Chapter 1 provides a review of the literature relevant to all chapters of the thesis. Chapters 2-5 are self-contained reports formatted in the style of a journal article and so each contains a review of the relevant literature, specific research aims, methods, results, and concise discussion. Finally, Chapter 6 provides a general discussion for the thesis, summarising the empirical chapters, situating the findings in the context of the general thesis aim, and highlighting new research questions that have arisen due to the thesis findings.

Thesis aim

This thesis is focused on impairments in the production and/or comprehension of language in people who have had a stroke or who have a neurodegenerative disease, termed post-stroke aphasia and primary progressive aphasia, respectively. Two clinical and theoretical issues are addressed in this thesis. First, despite the similarity of symptoms and nomenclature in subtypes of post-stroke aphasia (PSA) and primary progressive aphasia (PPA) (Grossman and Irwin, 2018), there have been few detailed direct comparisons across the full ranges of PSA and PPA. Secondly, although diagnostic subtypes have been proposed for both forms of aphasia, graded differences in aphasic profiles are evident between categories and there can be considerable within-subtype heterogeneity. Thus, the aim of the thesis was to compare directly the full ranges of PSA and PPA by exploring graded intragroup and intergroup differences.

To provide context for the clinical and theoretical issues which motivated this thesis, I will first provide an overview of post-stroke aphasia and primary progressive aphasia, placing the nosology of both forms of aphasia in the historical context which has led to their current classification systems. Second, I will summarise and evaluate previous comparisons of PSA and PPA. This will be followed by an overview of the multidimensional framework for reconceptualising patient variance along continuous dimensions. Finally, I will outline the specific aims of each empirical chapter.

The nosology of aphasia

The ability to use language is arguably one of the defining traits of human existence (Fitch, 2010). The focus of this thesis is a condition called aphasia, which arises when the ability to use language is affected by acquired brain injury. For the purpose of this thesis, aphasia will be defined as an impairment of the ability to comprehend and formulate language following acquired brain injury, which manifests as difficulties across multiple modalities of language use (e.g., reading, auditory comprehension, expressive language) (Rosenbek et al., 1989). This definition reflects the heterogeneity in aphasiological phenotypes which can be caused...
by brain injury. As is traditional in neurology and neuropsychology, aphasia presentations are classically grouped into subtypes with differentiating features (e.g., having preserved ability to repeat words and sentences vs. the loss of this ability). The current classification systems of post-stroke aphasia and primary progressive aphasia are summarised below, in the historical contexts which have influenced the nosology of these forms of aphasia (Tesak and Code, 2008).

Post-stroke aphasia

Aphasia is a common consequence of stroke in the left hemisphere (Berthier, 2005), affecting 30-38% of people in the acute setting after a first episode of stroke (Tsouli et al., 2009, Bersano et al., 2009). I will refer to this form of aphasia as post-stroke aphasia (PSA). Ischaemic stroke in a cerebral artery (e.g., middle cerebral artery) is a common cause of PSA, but other causes include haemorrhage or ischaemia in between vascular territories (Cauquil-Michon et al., 2011), or in subcortical brain areas (de Boissezon et al., 2005, Hillis et al., 2004, Wessels et al., 2006). People with PSA display considerable variation in the nature and severity of their impairments (e.g., naming, repetition, comprehension, reading, etc.) and consequently, as is traditional in neurology and neuropsychology, diagnostic classification systems have been developed to delineate subtypes of aphasia.

Classification of PSA into clinical phenotypes has its origins in the zeitgeist of 19th century research into the localisation of brain functions by Broca (Broca, 1861), Wernicke (Wernicke, 1874) and their contemporaries (Sondhaus and Finger, 1988, Ardila, 2010a, Berthier et al., 2014, Hillis, 2007). In the 20th century there was a shift of focus away from localisation of brain functions towards employing classification of aphasia to inform therapy, led by the ‘neoclassical’ approach of Geschwind which sought to integrate a therapy-focused approach to understanding aphasia with the revival of the localisationist perspective (Tesak and Code, 2008). Geschwind, Kaplan and colleagues developed the Boston Diagnostic Aphasia Examination (BDAE) (Kaplan, 1983). This classification system categorises patients with aphasia following a stroke into eight subtypes based on impairments in three language domains: fluency, ability to repeat, and ability to comprehend. Other diagnostic classification systems were being developed by contemporaries of Geschwind, such as the Western Aphasia Battery (WAB) (Kertesz, 1982) which also categorises patients into discrete subtypes and provides the Aphasia Quotient to give a sense of general aphasia severity regardless of subtype.
In the 1980s there were significant advancements in the ways to study the function and structure of the brain relating to language, including functional magnetic resonance imaging (Hillis, 2007). Progress since then has culminated in the current approach to aphasia classification which considers PSA presentations as vascular syndromes (Tippett and Hillis, 2016), which are constrained by the nature of the vascular territories that are affected in stroke. In fact, many of the founding ideas of the neo-associationist approach behind the taxonomies developed in the 1980s have since been proven inaccurate (Tremblay and Dick, 2016). Thus, the current approach to classification of PSA phenotypes appears to have evolved from the models and theories that drove the initial development of the BDAE and WAB and the connectionist theories in the 18th century. Despite this improved understanding, the taxonomic labels derived from the 18th century models remain pervasive in the literature (Kasselimis et al., 2017). Assuming contemporary researchers and clinicians are not using the diagnostic subtype labels to refer to the same aphasic profiles as those used to initially generate the labels in the 1980s, it may be concluded that there has been a natural evolution in the use of the subtype labels to parallel the greater understanding of the nature of variation in the patients’ presentations which is being driven by advancements in neuroimaging, computational modelling and other sophisticated methods to study aphasia.

**Primary progressive aphasia**

Another key factor influencing the change in the zeitgeist in the 1980s about classification of aphasia was the recognition of aphasia caused by neurodegenerative disease, because these forms of aphasia were associated with neural correlates outside of the regions typically affected by stroke (Hillis, 2007). However, progressive aphasia due to neurodegeneration was first described in the 19th century in the context of ‘local’ dementias impacting relatively selective aspects of cognition (Pick, 1892, Serieux, 1893, Déjerine and Séreix, 1897). Nearly a century later, aphasia caused by neurodegeneration which presented with fluent speech but impoverished vocabulary due to selective impairments in semantic memory was reported by Warrington et al. (1975). In 1982, Mesulam published a report on six cases of what he termed “slowly progressive aphasia without generalised dementia” (Mesulam, 1982). Aside from the uniting feature of a gradual, principal language disorder, the cases were heterogeneous; fluent and non-fluent aphasias were apparent, with and without word-comprehension difficulties. The fluent presentation of progressive aphasia described by Warrington et al. (1975) was later termed semantic dementia (Snowden et al., 1989). Concurrently, non-fluent aphasia due to neurodegeneration was investigated by research groups who focused separately on apraxia of speech (Turner et al., 1996) and agrammatism (Grossman et al., 1996) (i.e., deficits in the use of ‘function’ words and inflections (Kean, 2013)). Fluent and non-fluent progressive aphasias were later unified by Neary et al. (1998) under the labels of semantic dementia (SD) and progressive non-fluent aphasia (PNFA) as presentations of fronto-temporal lobar degeneration.
Primary progressive aphasia (PPA) is now the collective term used to refer to a group of dementias characterised by chronic and gradually progressing language impairment as a result of neurodegenerative disease, with the relative sparing of other domains of cognition and behaviour (Mesulam, 2001a). Unification of the varied forms of PPA as presentations of a predominant language disorder was proposed by international consensus (Gorno-Tempini et al., 2011). A two-stage diagnostic process was proposed, comprising criteria for neuropsychological and linguistic symptoms, neuroimaging markers and pathological signs. The first stage proposed criteria to give a root-diagnosis of PPA, and the second stage proposed criteria to classify into a PPA variant. The criterion for root-diagnosis of PPA is two years with progressive language impairments as the most salient symptom of a neurodegenerative disorder (Mesulam, 2001a). The consensus includes the three most commonly identified variants of PPA, though numerous additional subtypes are often proposed (such as agrammatic PPA without apraxia of speech (Tetzloff et al., 2019), or primary progressive apraxia of speech (Josephs et al., 2012)). The three variants in the consensus are outlined below.

The core features of non-fluent variant PPA (also known as progressive non-fluent aphasia, PNFA) are agrammatism in language production (i.e., shortened length of phrases, simplified syntactic structure, or problems with verbs or plurals (Kean, 2013, Turkstra and Thompson, 2011), and effortful speech (Gorno-Tempini et al., 2011). Atrophy and hypometabolism in PNFA are typical found in left posterior fronto-insular regions (Josephs et al., 2006, Gorno-Tempini et al., 2004, Nestor et al., 2003). PNFA is often associated with fronto-temporal lobar degeneration pathology, positive for micro-tubule associated protein tau (FTLD-tau) (Josephs et al., 2006, Knibb et al., 2006).

The semantic variant of PPA (also known as semantic dementia (SD), see below) is characterised by two core features of anomia (i.e., word retrieval difficulty (Maher and Raymer, 2004)) (Woollams et al., 2008) and impaired single-word comprehension (Gorno-Tempini et al., 2011, Snowden et al., 1989). Reading and writing are characterised by surface dyslexia (Woollams et al., 2007) and dysgraphia. The neuroimaging correlates of svPPA are predominant anterior temporal lobe atrophy, hypoperfusion or hypometabolism as seen with MRI, SPECT and PET (Mummery et al., 2000, Acosta-Cabronero et al., 2011, Mion et al., 2010, Lambon Ralph et al., 2017). The neuropathological correlate of svPPA is FTLD pathology positive for ubiquitin (FTLD-U) (Davies et al., 2005, Leyton et al., 2016a).

There is some controversy regarding the distinction between what has been described as semantic dementia and semantic variant PPA. The distinction is made on the grounds that in svPPA, there is a predominant impairment for comprehension of language stimuli and left-hemisphere atrophy in the perisylvian language areas, whereas in semantic dementia there is bilateral (though left-dominant) atrophy of anterior temporal lobes associated with impairment of comprehension of stimuli beyond the language domain. Whilst the patients’ pronounced anomia and verbal comprehension deficit are often prominent clinical features in
the initial clinical presentation (Mesulam, 2001b, Mesulam et al., 2013), careful evaluation invariably identifies a range of nonverbal comprehension deficits (Bozeat et al., 2002, Bozeat et al., 2003, Adlam et al., 2006b, Goll et al., 2010, Piwnica-Worms et al., 2010) even in early cases (Bozeat et al., 2000).

For example, Adlam et al. (2006a) showed that in seven patients meeting the criteria for fluent PPA (before the 2011 consensus criteria labelling this as semantic-variant PPA), they demonstrated impairments in both verbal and non-verbal semantic memory tests. Furthermore, these impairments were related to grey matter atrophy which was found in the temporal lobes bilaterally. This thesis will therefore refer to this presentation of PPA as semantic dementia – since this encapsulates all the deficits associated with temporal lobe atrophy which is seen in what is called svPPA and SD.

The core features of the logopenic variant of PPA (logopenic progressive aphasia, LPA) are impaired word retrieval and sentence repetition (Gorno-Tempini et al., 2011, Gorno-Tempini et al., 2008, Rohrer et al., 2010). Another distinguishing feature of LPA is that phonological paraphasias, usually sound substitutions, are well-articulated without distortions (Wilson et al., 2010). Single-word comprehension and confrontation naming are relatively spared in LPA (Gorno-Tempini et al., 2004). The atrophy, hypoperfusion and hypometabolism have been localised to the left posterior perisylvian and temporoparietal cortices in LPA (Rohrer et al., 2013, Leyton et al., 2012). The most common underlying pathology for LPA is Alzheimer’s disease (Leyton et al., 2016b, Knibb et al., 2006).

Comparisons of aphasia across aetiologies

As described above, the classification systems for PSA and PPA share similar differential features, for example the distinction between fluent and non-fluent aphasia, or differentiation based on impairment in semantic knowledge. Broadly speaking in terms of the characteristic impairments associated with the diagnostic categories, PNFA might be compared to Broca’s aphasia, LPA to Conduction aphasia, and SD to Wernicke’s aphasia (Grossman and Irwin, 2018). However, despite superficial behavioural similarities, detailed direct comparisons between PSA and PPA are rare. Therefore, it is still unclear if the degree and nature of the symptoms are the same, or if the vocabulary terms used to describe the patients and their symptoms are truly equivalent. The small number of previous comparative studies have been focused on either specific tasks or linguistic/cognitive domains.

The typical picture of post-stroke non-fluent aphasia is described as effortful, dysfluent speech coupled with phonological and syntactic impairments (Berthier et al., 2014), and these are also the patterns of impairment which are most commonly observed with progressive non-fluent aphasia (Mesulam, 2001a). It may seem, on the surface that these are two syndromes differing only in the underlying aetiology and thus the time-course of the condition. Patterson et al. (2006a) compared speech production and phonological deficits in a selection of non-fluent subtypes of PSA and PPA. They found that patients with
progressive non-fluent aphasia exhibited speech impairments which were more sensitive to the nature of the task, compared to post-stroke non-fluent aphasics. Furthermore, progressive non-fluent aphasics displayed more accurate reading aloud of single-words with exceptional spelling when they were presented in isolation compared to their presentation in a paragraph. Post-stroke non fluent patients produced similar levels of accuracy for reading the exceptional words whether isolated or in the paragraph. This highlights the fact that speech production in progressive non-fluent aphasia is especially impaired for connected speech, particularly if self-generated. The evidence presented by Patterson et al. (2006a) shows that simply being on the non-fluent side of the diagnostic classification systems for PSA and PPA does not necessitate similarity in the presentation of aphasia.

Jefferies and Lambon Ralph (2006) (see also Jefferies et al., 2008) compared semantically-impaired PSA and PPA patients on a range of linguistic and non-linguistic semantic tasks. They found that semantic impairments in SD and comprehension-impaired PSA reflect the degradation of different contributing aspects of cognition and also involve brain injury to different regions; the profile of impairment in the SD cases was indicative of a 'storage' disorder (Jefferies et al., 2009), wherein core semantic representations are degraded and no longer available for recall in any modality, and this corresponded with atrophy in the bilateral anterior temporal lobes. In contrast, the profile of impairment in the comprehension-impaired PSA patients suggested a deficit in accessing and working with retained semantic knowledge (Thompson et al., 2015), at least partly caused by impairment in executive function, and this was associated with their lesions in left inferior/temporoparietal cortex. So although impaired comprehension is a useful feature for delineating these two conditions from other forms of aphasia, it is clear that the comprehension deficits involved are not equivalent.

Thompson et al. (2013, 2012) compared subtypes of PSA and PPA based on their core deficits; Broca’s PSA was contrasted with PNFA as forms of aphasia displaying agrammatism whilst anomic PSA was contrasted with LPA as forms of aphasia with predominant word-finding difficulty. Using the Northwestern Naming Battery, they found overlapping impairments in agrammatic presentations of PSA and PPA, specifically in producing verbs compared to nouns. No differences in verb/noun production were found for the anomic presentations (LPA and Anomic PSA). Using the Northwestern Assessment of Verbs and Sentences, they confirmed the above finding for verb deficits in agrammatic forms of PSA and PPA. They also found that PNFA and Broca’s aphasia showed overlapping impairments in comprehension and production of syntactically complex sentences. Their results suggest that agrammatism results in impaired processing of complex syntax regardless of the aetiology of the brain injury causing the agrammatism.

These key studies represent an important step towards understanding aphasia transdiagnostically. One line of investigation that is missing from the above comparisons is directly comparing the neural correlates of aphasia across aetiologies; PSA and PPA
represent complementary perspectives for understanding aphasia because there are key
differences in the location and nature of brain injuries that cause PSA and PPA (Grossman
and Irwin, 2018). These include whether language regions are affected unilaterally (usually
left hemisphere in PSA (Berthier, 2005)) or bilaterally/left-dominantly (Acosta-Cabronero
et al., 2011), the location of regions affected (dictated by vascular territory (Hillis, 2007) or the
spread of pathology (Leyton et al., 2016a)), and whether white matter structures (Ho et al.,
2005, Seeley et al., 2009) or specific cortical layers (Romito-DiGiacomo et al., 2007) are
principally affected. Thus, studying neural correlates of PSA or PPA in isolation means
investigating only certain aspects of aphasia, as limited by the nature, timing, and location of
the neural damage. Investigating the neural correlates of aphasia transdiagnostically, in PSA
and PPA concurrently, means examining more presentations of aphasia due to more
diversity in the nature, timing, and location of the neural damage.

Comparisons of PSA and PPA which consider structural brain abnormalities are scarce, and
few have compared PSA and PPA brain scans in the same neuroimaging analysis. Ogar et
al. (2011) found that Wernicke’s aphasia was associated with greater impairment on
spontaneous speech and sentence comprehension measures compared to SD. To relate
these behavioural differences to the different patterns of brain injury, the region of greatest
neural damage in SD (left anterior temporal lobe) was established using voxel-based
morphometry, whilst the lesions in Wernicke’s aphasia (covering left posterior middle
temporal gyrus) were demarcated by hand. Measuring the extent of neural damage
differently in each group introduced a confounding variable which means that the
relationship between impairment and neural damage could only be directly assessed within
each group; the extent of damage could not be compared directly across aetiologies since
the brain scans were not part of the same analysis procedure.

Budd et al. (2010) compared naming errors in PPA and acute PSA to establish whether
there were differences due to the aetiology or the location of the brain damage. Acute PSA
cases were categorised based on the location of their stroke (middle cerebral artery (either
anterior, posterior or mixed), posterior cerebral artery, anterior cerebral artery, or
subcortical). PPA cases were classified into a subtype (semantic variant PPA, non-fluent
variant PPA, or logopenic variant PPA) using clinical and imaging criteria (Gorno-Tempini et
al., 2011). In terms of the location of neural damage, non-fluent variant PPA was matched
with anterior middle cerebral artery stroke, whilst logopenic variant PPA was matched with
posterior middle cerebral artery stroke. Semantic variant PPA did not have a form of stroke
with a corresponding location of neural damage. When naming performance was compared
across location of damage, there were no significant differences between aetiologies.
However, when naming performance was compared across aetiology, there were
significantly fewer circumlocutions in the anterior middle cerebral artery (MCA)/anterior
perisylvian cortex group than the other locations of damage. This study highlights the utility
of a transdiagnostic approach to understanding the neural correlates of language
impairments. However, as noted by the authors, the location of brain damage was coarsely defined and the neuroimaging analysis of the PSA and PPA brains in the same model was not leveraged to compare the groups.

Lastly, Silveri et al. (2019) assessed language impairments in ‘fluent’ forms of PSA and PPA. Specifically, and exceptionally, they defined their fluent PPA cases as logopenic variant/mixed primary progressive aphasia (lvmPPA). This categorisation represents a response to previous research which has highlighted (a) that lvPPA is diagnosed by excluding the other forms of PPA, and (b) that lvPPA show highly overlapping symptoms and neural correlates with cases of PPA which fail to meet the criteria for a subtype of PPA and thus are labelled ‘mixed’ PPA (mPPA) (Sajjadi et al., 2012a, Sajjadi et al., 2014, Silveri et al., 2014). The lvmPPA cases in Silveri et al. (2019) were matched with a fluent presentation of PSA consistent with Wernicke’s aphasia or conduction aphasia. Regions of significant brain injury were confirmed in lvmPPA using cortical thickness measurements, whilst lesions in PSA were manually demarcated using MRIcron. Scores derived from principal component analysis (PCA) on four extracted components (reflecting aspects of morphosyntactic and lexical processing) were used in a cluster analysis to compare profiles of lvmPPA and fluent PSA. The first cluster contained mostly lvmPPA (78%) cases, whilst the second cluster contained mostly PSA (64%) cases. lvmPPA cases in the first cluster had more bilateral cortical thinning than lvmPPA in the second cluster. The authors reported no obvious differences in the lesions of PSA cases in the two clusters. Overall, this study showed that location of atrophy was important for the clusters of lvmPPA which were generated based on impairments in language processing. However, the relationship between forms of fluent aphasia and location of brain damage could not be compared directly across lvmPPA and PSA due to the different approaches to quantify neural injury.

The previous comparisons of PSA and PPA have highlighted important similarities and differences in aphasia across these aetiologies. However, larger scale comparisons are required in order to (a) simultaneously compare a broad range of linguistic/cognitive symptoms and situate relative impairments in the same symptom space, and (b) create a framework which can incorporate all presentations of aphasia, including those which are often excluded from comparisons based on subtypes (i.e., those with ‘mixed’ aphasia).

With regards to comparing the full spectrum of language and cognitive symptoms in PSA and PPA, a key issue is the lack of a neuropsychological test battery which has been validated for PPA, let alone validated for both PSA and PPA. In contrast, there are many language assessment tools which are often used for (i.e., have been validated for) PSA (e.g., Boston Diagnostic Aphasia Examination (Kaplan, 1983), Western Aphasia Battery (Kertesz, 1982), Comprehensive Aphasia Test (Swinburn et al., 2004), Aachen Aphasia Test (Huber et al., 1984), Quick Aphasia Battery (Wilson et al., 2018)). These could be considered to be ‘established’ tools (or even ‘gold standard’ tools), as they have been extensively validated and are used and trusted by researchers and clinicians alike for
assessing PSA. Some of the above assessment tools have been applied to assess language impairments in PPA. For example, the WAB has been applied to PPA to quantify the severity of language impairment through the aphasia quotient (Catani et al., 2013; Mesulam et al., 2009; Tetzloff et al., 2019; Thompson et al., 2012; Utianski et al., 2019) and to compare linguistic impairments in PPA with AD (Kertesz et al., 2003) and behavioural-variant frontotemporal dementia (Blair et al., 2007). The Comprehensive Aphasia Test has also been used to monitor response to therapy in single cases of PPA (Hameister et al., 2017; Taylor et al., 2014).

However, the applicability of the BDAE or WAB subtype categories to the presentations of PPA is called into question by comparison studies outlined above which show subtle differences even in superficially similar classifications. Furthermore, the current core characteristics for differential diagnosis of PPA subtypes (Gorno-Tempini et al., 2011), are not all assessed in the established tools for PSA (e.g., the WAB does not assess reading, writing, sentence comprehension with pictures, non-word repetition, or semantic association). Thus, there is a need for a tailored language assessment tool for PPA.

The Mini-Linguistic State Examination (MLSE) is a novel clinical assessment tool being developed to specifically assess the major linguistic domains commonly affected in primary progressive aphasia (Gorno-Tempini et al., 2011). Additionally, as part of a drive to increase comparisons of language impairments across aetiologies of aphasia, the MLSE is being designed to have equivalent validity for assessing language impairments caused by stroke, or language impairments which form a concomitant symptom in other neurodegenerative disorders. For example, corticobasal syndrome (CBS) and progressive nuclear palsy (PSP) are classified as movement disorders which are part of the spectrum of frontotemporal lobar degeneration (FTLD), which also contains PNFA and SD. Although classically conceptualised as disorders affecting movement, CBS and PSP also show language impairments (Burrell et al., 2018, Kertesz et al., 2000). However, differences in assessment of impairments in studies of CBS and PSP hinder conclusions about characteristic language symptoms in these forms of FTLD (for review see Peterson et al., 2019). In general, language impairments in PSP are consistent with a frontally-mediated impairment (Rosser and Hodges, 1994) in executive function (Magherini and Litvan, 2005). CBS is associated with a breakdown in language processing resulting in recognisable aphasic phenotypes such as Broca’s aphasia (Frattali et al., 2000). To address the lack of consensus regarding language impairments in CBS and PSP, the MLSE will provide a vital platform for increased, consistent assessment of language deficits in these forms of FTLD.

As the MLSE is a novel tool for PPA, there are no equivalent tests in PPA which the MLSE can be validated against. Therefore, the MLSE must be validated in comparison with an established tool (Ivanova and Hallowell, 2013) which has been developed for PSA, such as the BDAE (Kaplan, 1983). Since such tools have been validated for PSA, the first formal evaluation of the MLSE must use a cohort of PSA to compare how the deficits captured by...
the MLSE compare to those captured by the established tool. Therefore, one of the empirical chapters of this thesis aimed to validate the MLSE against the BDAE in a cohort of chronic PSA. The validity of the MLSE was assessed in terms of convergent validity (Mitrushina et al., 2005), sensitivity to mild deficits and sensitivity to levels of impairment. These features are important for a clinical assessment tool for a number of reasons including ensuring that individuals with mild aphasia are not mistakenly discharged/withdrawn from clinical services due to lack of apparent impairment (Johnson et al., 1998), capturing an accurate representation of the epidemiology of aphasia due to stroke requires identifying the full spectrum of language impairments (Flowers et al., 2016), and informing interactions with clinical pathways (Shrubsole et al., 2017) such as speech and language therapy.

In conclusion, comparisons of language impairments in PSA, PPA and potentially in movement disorders, represent a unique opportunity to leverage the differences between these conditions to better understand the core language systems that are affected in aphasia regardless of aetiology. Large-scale comparisons of the full ranges of PSA and PPA (including ‘mixed’ cases, see below) are needed to avoid the pitfalls of comparing cases based on their diagnostic category membership. Furthermore, comparisons should make use of assessment tools/test batteries which measure a diverse range of linguistic symptoms. By addressing these considerations, the empirical studies in the thesis aimed to capture the heterogeneity within and between the clinical phenotypes of these forms of aphasia.

Reconceptualising variance along multiple phenotypic spectra

There is considerable heterogeneity in the nature and severity of language symptoms displayed by people with PSA and PPA. Conceptually, this variation is treated as arising due to an underlying structure in aphasia, rather than attributed to random variation and noise. Underlying structures of variance in health and disease fall into two types, namely categorical variation or continuous variation. In the context of aphasia, a categorical variance structure implies that aphasia phenotypes can be split into mutually exclusive coherent groups or clusters. Such categorical systems are based on two assumptions: 1) that there is homogeneity within each category; and 2) that there are distinct boundaries between categories (Schwartz, 1984). Alternatively, a continuous underlying structure of variance in aphasia assumes graded differences between aphasia phenotypes.

The classical diagnostic classification systems developed since the 1980s are associated with some important clinical and research benefits. For example, the BDAE classification system facilitated therapy programmes for Broca’s aphasia (Helm-Estabrooks and Ramsberger, 1986), and global aphasia (Helm-Estabrooks et al., 1982). However, others have argued against the utility of broad ‘brush stroke’ classifications like Broca’s aphasia for describing an individual’s unique constellation of impairments in the context of therapy (Feyereisen et al., 1986; Gordon, 1998). Classification systems provide a useful diagnostic short-hand for clinicians, and may be useful for contrastive group-level analysis (Zurif et al.,
Nevertheless, the categorical subtypes of aphasia delineated by classification systems like the BDAE are pervasively used in clinical settings and research settings. It is important to acknowledge explicitly here that clinicians/researchers who make use of the diagnostic categories do not hold the belief that the structure of variance in aphasia actually reflects mutually exclusive diagnostic categories (we are not attempting to build a ‘strawman’ to fight against (Kasselimis et al., 2017)). However, the traditional diagnostic classification systems outlined above, which are still pervasive in the literature (Kasselimis et al., 2017, Kasselimis et al., 2012), are inherently based on the assumption that variance in aphasia is (at least somewhat) categorical. This is because the categories are defined by dichotomous signs (e.g., spared/impaired repetition, or impairment above/below a cut-off threshold) which result in symptoms which can be differentially diagnosed. There is evidence that this approach to conceptualising variance does not capture the full phenotypic space of aphasia. Firstly, a considerable proportion of people with PSA or PPA are classified as having ‘mixed’ aphasia because they do not fulfil the criteria for one diagnostic category (this could be due to fulfilling the criteria for more than one subtype or not fulfilling the criteria for any subtype) (Benson, 1979, Botha et al., 2015, Wertz et al., 1984, Wicklund et al., 2014, Sajjadi et al., 2012a, Harris et al., 2013, Matias-Guiu et al., 2014, Mesulam et al., 2012, Utianski et al., 2019, Gil-Navarro et al., 2013, Mesulam et al., 2008, Spinelli et al., 2017, Knibb et al., 2009).

In fact, the BDAE and WAB acknowledge that due to the heterogeneity in PSA (Butler et al., 2014), a large percentage of cases are unclassifiable (Kaplan, 1983, Kertesz, 2007) and the BDAE may fail to classify 30-80% of patients (Benson, 1979, Wertz et al., 1984). These mixed and unclassifiable cases suggest that there is variation amongst aphasia subtypes which cannot be accounted for in categorical classification system.

For example, many studies have found overlapping language impairments across different subtypes of PSA; Thompson et al. (2015) found that Wernicke’s aphasia and semantic post-stroke aphasia (defined as ‘a multimodal semantic deficit in which there is difficulty in the controlled ‘application’ of knowledge correlating with executive impairment’ (Thompson et al., 2015, pp. 3777)) show overlapping deficits in semantic access (accessing correct semantic knowledge). Robson et al. (2012) also found similar impairments in executive control of semantic processing in these subtypes of PSA. There is also heterogeneity within the proposed subtypes of PSA. For example, Anomic PSA can be caused by lesions in various cortical and subcortical left-hemisphere regions (Kertesz, 1979), meaning that patients classified as ‘Anomic’ are a highly heterogeneous group (Maher and Raymer, 2004).

In PPA, formal evaluations of the 2011 diagnostic criteria have highlighted cases which are poorly classified by the classification system (Harris et al., 2013, Wicklund et al., 2014, Sajjadi et al., 2012a). Other studies have suggested further subdivisions of the diagnostic
categories (e.g., PNFA might contain agrammatic and apraxic subtypes (Tetzloff et al., 2019)), highlighting heterogeneity within the aphasic presentations which are considered to fall into these diagnostic categories. Furthermore, Sajjadi et al. (2014, 2012a) have shown that cases of mixed PPA show a combined profile of deficits, atrophy, and pathology which reflects the characteristics of LPA and underlying Alzheimer’s disease. They conclude that the diagnostic criteria for LPA are unable to capture the heterogeneity in language impairments caused by Alzheimer’s disease.

Alzheimer’s disease pathology is known to cause a heterogeneous collection of clinical phenotypes, which current diagnostic criteria delineate into typical (i.e., amnestic (Varma et al., 1999)) and atypical (non-amnestic) presentations (Dubois et al., 2014). However, there is evidence that the AD diagnostic categories also fail to capture the full phenotypic space of typical and atypical AD. This evidence has generated the hypothesis that the relationship between variants of AD might reflect graded variation along one or multiple phenotypic spectra (Fitzpatrick et al., 2019; Lambon Ralph et al., 2003; Migliaccio et al., 2009; Ridgway et al., 2012; Stopford et al., 2008; Warren et al., 2012). Thus, there is a direct parallel between aphasia and AD, in terms of heterogeneous and gradedly-different subtype profiles and accompanying drawbacks of categorical classification systems.

Typical AD is defined by impairment in hippocampal-dependent episodic memory (Varma et al., 1999). Atypical AD is characterised by relatively preserved episodic memory in comparison with impairments in another cognitive domain (Dubois et al., 2014), such as executive function, linguistic processing, or visual processing (Dubois et al., 2007). However, there are two key observations which suggest that the categorical classification system for variants of AD fails to capture the true phenotypic space in these disorders: (1) there is considerable overlap in the key features (behavioural and neural) of typical and atypical AD, and (2) there is considerable heterogeneity within the proposed category of typical AD.

Studies comparing behavioural profiles across variants of AD have found considerable overlap in key behavioural features within the AD variants. For example, Crutch et al. (2012b) found overlapping linguistic and praxic deficits in posterior cortical atrophy and logopenic progressive aphasia. Furthermore, Migliaccio et al. (2009) also found overlapping clinical features in posterior cortical atrophy, logopenic progressive aphasia and early-onset AD, with greater overlap as the disease progressed in each group. These overlapping features suggest blurred between-group boundaries, as is the case between subtypes of aphasia. Heterogeneity within the typical AD variant has also been demonstrated, with many studies showing that the characteristics of atypical AD are echoed within typical AD (Peter et al., 2014, Caine and Hodges, 2001, Kanne et al., 1998, Price et al., 1993, Snowden et al., 2007, Stopford et al., 2008); for example, Lambon Ralph et al. (2003) found cases of typical AD who displayed impairments in verbal processing or visuospatial processing, i.e., resembled atypical-like presentations of typical AD. These findings suggest that typical AD
and atypical AD might be better reconceptualised as representing parts along a continuous phenotypic spectrum.

This reconceptualisation is relevant for the parietal atypical variant of AD. This variant is also known as posterior cortical atrophy (Benson et al., 1988), and the core symptoms are elements of Bálint’s syndrome (such as optic ataxia, simultanagnosia, oculomotor apraxia) and elements of Gerstmann’s syndrome (such as acalculia, agraphia, finger agnosia and left-right disorientation) (Crutch et al., 2017, Crutch et al., 2012a). Mirroring the ventral/dorsal routes of healthy visual processing (Goodale and Milner, 1992), ventral and dorsal subtypes of posterior cortical atrophy have been described (Migliaccio et al., 2012, Tang-Wai et al., 2004, Tsai et al., 2011, Ross et al., 1996). However, others have reported overlap between ventral/dorsal groups in atrophy patterns and symptoms (Crutch et al., 2017, Lehmann et al., 2011a), leading to the hypothesis that these presentations of posterior cortical atrophy represent different ends of a spectrum of phenotypic variation.

To summarise the studies of PSA, PPA, AD and posterior cortical atrophy described in this section, patients with these neurological and neurodegenerative disorders display considerable variation in the nature and severity of their impairments and this variation is not always adequately captured by the categorical classification systems in use. Thus, the hypothesis which is addressed in this thesis is that variation in patients’ phenotypes may be best conceptualised using an underlying continuous structure, i.e., comprising multiple graded dimensions. This hypothesis is motivated by recent studies which have employed formal explorations of the graded nature of variance in aphasia (Butler et al., 2014; Mirman et al., 2015a). In these studies, the underlying structure of variance is conceptualised as a multidimensional ‘space’, the axes of which represent principal language-cognitive domains that are affected across aphasia phenotypes. Many studies employing this alternative, non-categorical approach have found that the axes of the multidimensional space reflect variation in phonology, semantics, speech fluency, and (if assessed) executive function (Butler et al., 2014; Halai et al., 2017; Halai et al., 2018a; Schumacher et al., 2019). Thus, patients’ individual aphasic profiles can be situated within the multidimensional space by using their position along the various axes (e.g., phonology) as coordinates. Using this multidimensional framework, graded differences between cases with different categorical diagnostic labels (e.g., Broca’s aphasia or Global aphasia) can be visualised in the overlapping regions of the space which are occupied by these cases.

By way of analogy, one can think of patients as colour hues across the red, green and blue (RGB) colour space. Each dimension of this colour space is defined by variation in a single colour. Whilst it is possible to demarcate and label (cf. categorise) approximate areas in the space as yellow (e.g., Broca), blue (Wernicke), etc., there are in fact many different kinds of each colour and the boundaries between them are fuzzy. Likewise, when presented with individual hues it is not always obvious which colour category they fall into (e.g., teal, maroon, indigo; cf. how to categorise a patient with mixed aphasia). Thus, like aphasia
classifications, colour labels provide approximate information about the underlying graded differences. This is sufficient to communicate broad distinctions between cases (e.g., blue vs. yellow; Broca vs. Wernicke) but not finer variations (the overlapping variations of orange vs. yellow; conduction vs. Wernicke). An alternative and more precise approach is to represent each hue (patient) in terms of its position along the RGB dimensions or coordinates within the multidimensional space (cf. describing patients' performance in terms of the underlying primary language-cognitive systems).

Principal component analysis (PCA) is an exploratory, data-driven analysis method which can be used to capture the principal dimensions of variance which, in the proposed framework above, represent the axes of the multidimensional space. PCA is a form of factor analysis which extracts a set of linearly uncorrelated variables (called components or factors – these terms are used interchangeably to refer to the variables which are created by PCA) from a set of inter-correlated observations (Jain and Shandliya, 2013). With respect to the colour analogy above, the components extracted by PCA are analogous to the colour axes.

Every variable (e.g., test) entered into the PCA is assigned a loading (called a factor loading) onto each of the extracted components. Individual cases (e.g., patients) are assigned scores on the extracted components (called factor scores). These factor scores represent a weighted average score of a patient's raw scores over all tests, weighted by the PCA-derived coefficient for each component. To aid cognitive interpretability of the extracted components, varimax rotation can be used to simplify the sub-space generated by the PCA. This rotation transforms the component-coordinate system in order to make the components orthogonal, i.e., statistically independent, with the result of the varimax rotation being that factor loadings can be explained more economically, and the principal components can be interpreted more easily (Kaiser, 1958).

Patients' factor scores can be used as coordinates on the language-cognitive axes of the multidimensional space to reflect where patients 'sit' in the multidimensional space (cf. how the coordinates of regions in the RGB colour space can be described by relative amounts of each colour). In this framework, the regions of the multidimensional space which are sampled by (i.e., occupied by) different diagnostic categories can be visualised graphically by plotting patients' factor scores within the multidimensional space of the principal components. Visualising variance in this way (Butler et al., 2014, Mesulam et al., 2009, Armstrong et al., 2000, Martin et al., 1986) can highlight whether it appears to conform to a categorical underlying structure or a continuous underlying structure. This method therefore represents a key way to approach the conceptualisation of graded variance in aphasia or AD.

Furthermore, by extracting orthogonal principal components PCA generates variables which are ideal for implementation into an analysis with voxel-based correlational methodology (VBCM) (Tyler et al., 2005). This is a variant of voxel-lesion symptom mapping (VSLM)
Unlike VLSM, VBCM does not require a binary classification of the intact/abnormal neural tissue because both the behaviour and tissue integrity measures are treated as continuous variables. By combining the continuous voxel intensity values with the continuous, orthogonal PCA component scores, the continuous nature of neural structures and the graded behavioural variation of patients can be preserved. This method allows the unique neural correlates of the orthogonal components to be isolated, thereby circumventing the issue of overlapping brain areas identified due to test co-linearity. Previous groups have used this PCA and VBCM approach to identify the unique neural correlates of principal language dimensions across subtypes of PSA (Butler et al., 2014, Halai et al., 2017, Halai et al., 2018a, Halai et al., 2018b, Schumacher et al., 2019, Woollams et al., 2018), but it has not been applied to PPA, or used as a platform for comparing PSA and PPA.

In conclusion, this multidimensional approach using PCA and VBCM has potential for uncovering brain-behaviour relationships of language impairments across different aetiologies of aphasia. This has yet to be explored in studies which (a) compare directly the full phenotypic spectra of aphasic presentations across aetiologies, (b) use a comprehensive test battery to simultaneously assess a broad range of linguistic symptoms, (c) explore graded variation within and between proposed categorical subtypes of PSA and PPA in a framework which can account for mixed presentations, and (d) compare brain injury metrics due to stroke and neurodegeneration in the same neuroimaging analysis models to relate these to shared neural correlates across aetiologies of aphasia. This important gap is addressed by the studies presented in this thesis.

Specific aims per chapter

A key overall aim of the thesis was to compare directly the full ranges of PSA and PPA using a graded, multidimensional framework to map out graded intragroup and intergroup differences. Importantly, the aim of situating PSA and PPA in a shared multidimensional space was not to differentiate between these aetiologies. Rather, the shared space is used as a platform for the large-scale direct comparison without selecting specific subtypes or cognitive/linguistic processes. This was addressed with the first three empirical chapters: Chapter 2 involved a behavioural comparison of PSA and PPA using the graded multidimensional framework; Chapter 3 involved validation of a novel linguistic assessment tool to use as a basis for comparing aphasia across aetiologies in Chapter 4; Chapter 4 extended the results of Chapter 2 by including exploration of the neural correlates of the dimensions found in Chapter 2. A second overall aim of the thesis was to explore the underlying structure of patient variance using a multidimensional framework. This was addressed in Chapters 2 and 4 which situated PSA, PPA and movement disorders in a shared aphasiological multidimensional space, and also in Chapter 5 which extended this framework to variation in typical and atypical AD.

The specific aims per chapter were:
1. To compare behavioural variation in PSA and PPA using PCA. Specifically, to compare the full ranges of PSA and PPA on a substantial linguistic and cognitive behavioural battery, using PCA to explore graded differences along the unique and shared dimensions extracted. Finding graded or categorical differences, to formally quantify how category-like each proposed subtype of PSA or PPA is within the aetiology-specific multidimensional space.

2. To validate a novel clinical tool for assessing aphasia across aetiologies. To conduct the first formal validation of the Mini-linguistic State Examination (MLSE) in order to establish (1) the convergence, or lack of convergence, of the subtype classifications derived from MLSE test scores vs. test scores on an established aphasia test; (2) sensitivity of the MLSE compared to the in-depth tests in terms of (a) sensitivity to mild deficits, and (b) sensitivity to levels of impairment; and (3) convergent validity of the MLSE subtests compared with the in-depth tests.

3. To compare behavioural and neural variation in language impairments across aetiologies. Specifically, to compare language impairments on the same transdiagnostic assessment tool (MLSE), to situate these impairments in a shared, multidimensional space using principal component analysis, then to uncover shared neural correlates associated with the behavioural dimensions, by comparing neural abnormality across aetiologies in the same voxel-based morphometry analysis.

4. To investigate the generalisability of the graded, multidimensional framework to variation in Alzheimer’s disease. Specifically, to explore graded behavioural variation in typical and atypical Alzheimer’s disease, to evaluate whether this relationship is comprised of categorical differences (implying discrete subtypes) or graded variation (implying one or multiple phenotypic spectra).
Chapter 2 - Graded, multidimensional intragroup and intergroup variations in primary progressive aphasia and post-stroke aphasia

Ruth U. Ingram¹, Ajay D. Halai², Gorana Pobric¹, Seyed Sajjadi³,
Karalyn Patterson²,⁴ & Matthew A. Lambon Ralph²

¹Division of Neuroscience and Experimental Psychology,
School of Biological Sciences, University of Manchester, UK
²MRC Cognition & Brain Sciences Unit, University of Cambridge, UK
³Department of Neurology, University of California, Irvine, Irvine, USA
⁴Department of Clinical Neurosciences, University of Cambridge

Statement of contribution
Ruth Ingram analysed all the data presented in this chapter. Seyed Sajjadi and Karalyn Patterson collected the behavioural data in this chapter. Matt Lambon Ralph, Ajay Halai, and Gorana Pobric provided guidance and support throughout the analysis, and feedback on drafts of the write up.
Abstract

Language impairments caused by stroke (post-stroke aphasia) and neurodegeneration (primary progressive aphasia) have overlapping symptomatology, nomenclature and are classically divided into categorical subtypes. Surprisingly, primary progressive aphasia (PPA) and post-stroke aphasia (PSA) have rarely been directly compared in detail. Rather previous studies have compared certain subtypes (e.g., semantic variants) or have focussed on a specific cognitive/linguistic task (e.g., reading). This study assessed a large range of linguistic and cognitive tasks across the full spectra of PSA and PPA. We applied varimax-rotated principal component analysis to explore the underlying structure of the variance in the assessment scores. Similar phonological, semantic and fluency-related components were found for PSA and PPA. A combined principal component analysis across the two aetiologies revealed graded intragroup and intergroup variations on all four extracted components. Classification analysis was employed to test, formally, whether there were any categorical boundaries for any subtypes of PPA or PSA. Semantic dementia proved to form a true diagnostic category (i.e., within group homogeneity and distinct between group differences), whereas there was considerable overlap and graded variations within and between other subtypes of PPA and PSA. These results suggest that (a) a multidimensional rather than categorical classification system may be a better conceptualisation of aphasia from both causes, and (b) despite the different aetiologies of pathology, these broad classes of aphasia have considerable features in common.
Introduction

Aphasia is an impairment of the ability to comprehend and formulate language following acquired brain damage, which manifests as difficulties across multiple modalities of language use (e.g., reading, auditory comprehension, expressive language) (Rosenbek et al., 1989). Causes of acquired brain damage leading to aphasia include stroke and neurodegeneration. The latter cause results in a form of aphasia termed primary progressive aphasia (PPA) (Mesulam, 2001a). To differentiate the two, we will refer to aphasia as a consequence of stroke as post-stroke aphasia (PSA). Two clinical and theoretical issues are addressed in this study. Firstly, despite the similarity of symptoms and nomenclature in subtypes of PSA and PPA, there have been few – if any – detailed direct comparisons across the full ranges of PSA and PPA. Secondly, although diagnostic subtypes have been proposed for both forms of aphasia, patients often vary greatly within each category or commonly fall between classifications (and thus are referred to as ‘mixed’). This suggests that the phenotype differences observed across patients might reflect graded variations across multidimensional aphasic spectra rather than a series of mutually exclusive, coherent diagnostic categories (Lambon Ralph et al., 2003, Warren et al., 2012, Stopford et al., 2008, Migliaccio et al., 2009, Ridgway et al., 2012). By combining detailed assessment data across the full ranges of PSA and PPA, this study was able to map out these graded intergroup and intragroup variations.

Although arising from different pathologies, PSA and PPA share symptomatology. Despite these clear superficial behavioural similarities, detailed direct comparisons between PSA and PPA are rare and thus it is still unclear, if the degree and nature of the symptoms are the same, or if the vocabulary terms used to describe the patients and their symptoms are truly equivalent. The small number of previous comparative studies have been focused on either specific tasks or linguistic/cognitive domains. For example, Patterson et al. (2006a) compared speech production and phonological deficits in a selection of non-fluent subtypes of PSA and PPA. Jefferies and Lambon Ralph (2006) compared semantically-impaired PSA and PPA patients on a range of linguistic and non-linguistic semantic tasks (see also Jefferies et al., 2008). Thompson et al. (2013) compared syntactic processing in agrammatic and anomic forms of PSA and PPA (see also Budd et al., 2010, Faria et al., 2013, Thompson et al., 2012). Whilst these important studies have advanced our understanding of specific language features for selected subtypes of PPA/PSA, larger scale studies are needed for at least two reasons: (a) it is important to explore performance simultaneously across a broad spectrum of language and cognitive areas in order to situate and understand any one specific task; and (b) comparisons of select PPA/PSA subtypes make the assumption that the subtypes can be readily identified and are the most appropriate basis for the comparison.

People with PPA or PSA display considerable variation in the nature and severity of their impairments (e.g., naming, repetition, comprehension, reading, etc.) – but what is the basis
of these variations? To clinical and research professionals working with people with aphasia, it is clear that there is an underlying structure in their aphasic performance (i.e., heterogeneity in aphasia phenotype is not caused by random variation and noise). Ruling out random variation, behavioural variation in health or disease can be split into two types reflecting the presence of either multiple, mutually exclusive coherent categories of person/patient, or graded variations along different dimensions. All true categorical classification systems are based on two assumptions: 1) that there is homogeneity within each category or type; and 2) that there are distinct boundaries between categories (Schwartz, 1984).

As is traditional in neurology and neuropsychology, categorical subtypes of PSA and PPA have been proposed. The contemporary taxonomy of PSA can be described as a neo-associationist classification approach (Tesak and Code, 2008), from which several enduring classification systems have been developed and are widely used today. The Boston Diagnostic Aphasia Examination (BDAE) (Kaplan, 1983), for example, categorises PSA patients into one of seven subtypes based on their relative strengths and weaknesses in repetition, speech output fluency and comprehension. In addition to this, the BDAE also uses this information to give an indication of the level of impairment within each subtype. The Western Aphasia Battery (WAB) (Kertesz, 2007) categorises patients into discrete subtypes and provides the Aphasia Quotient to give a sense of general aphasia severity regardless of subtype. Likewise, the consensus derived classification system for PPA (Gorno-Tempini et al., 2011) delineates three categorical subtypes: semantic dementia/semantic variant PPA, non-fluent/agrammatic variant PPA and logopenic variant PPA, though numerous additional subtypes are often proposed (such as agrammatic PPA without apraxia of speech (Tetzloff et al., 2019), or primary progressive apraxia of speech (Josephs et al., 2012)).

There is evidence, however, that a strict categorical approach is limited and does not capture the true nature of the patients’ variations. Thus, (a) rather than homogeneity within each category, there is significant variation (e.g., the different presentations of non-fluent progressive aphasia (Tetzloff et al., 2019)); (b) patients’ categorical membership can change (with recovery in PSA and decline in PPA); and (c) there can be blurred boundaries between categories (e.g., Broca-like or Wernicke-like conduction aphasia (Song et al., 2011)). One consequence is that in both PSA and PPA there is a considerable proportion of patients who must be classified as having ‘mixed’ aphasia because they either do not fulfil the criteria for any subtype, or even fulfil the criteria for more than one subtype (Benson, 1979, Botha et al., 2015, Wertz et al., 1984, Wicklund et al., 2014, Sajjadi et al., 2012a, Harris et al., 2013, Matias-Guiu et al., 2014, Mesulam et al., 2012, Utianski et al., 2019, Gil-Navarro et al., 2013, Mesulam et al., 2008, Spinelli et al., 2017, Knibb et al., 2009). These limitations of the categorical approach implied by the syndrome classification systems have long been understood clinically (Caramazza, 1984) (e.g., the limited use of broad ‘brush stroke’
classifications like Broca's aphasia in describing an individual's unique impairments for the purpose of therapy (Feyereisen et al., 1986, Gordon, 1998)). Despite these drawbacks, 'the neo-associationist [sic] classification still haunts the literature' (Kasselimis et al., 2017, pp. 64). The question arises then, how to reconcile the increased awareness of heterogeneity between and within PSA subtypes, with the continued use of the diagnostic category labels?

The PSA subtypes derived in the neo-associationist approach are based on connectionist models which made use of then-contemporary knowledge of brain-behaviour relationships in PSA (Tesak and Code, 2008). However, our understanding of PSA has advanced significantly from these connectionist models, many of which have been shown to be inaccurate (Tremblay and Dick, 2016). Consequently, modern uses of the diagnostic labels derived from these now-updated models must either (a) conform to the original concepts of PSA subtypes (shown to be inaccurate in many cases), or (b) use the labels to refer to an updated understanding of PSA subtypes. We propose that, in PSA, the use of the subtype labels from the traditional, neo-associationist, classification systems has evolved over time to reflect the greater understanding of brain-behaviour relationships in PSA. Our overarching hypothesis is that patients with aphasia vary along graded dimensions, and we propose that the contemporary use of the diagnostic category labels has evolved to reflect this graded variation. This hypothesis arises from formal explorations of an alternative, non-categorical way to conceptualise behavioural variations (Butler et al., 2014, Mirman et al., 2015a).

These new approaches are based on the second source of individual differences noted above – namely, graded variations along continuous behavioural dimensions. Recent studies have reconceptualised the variations in PSA as forming an aphasic multidimensional space with each patient taking up a different position (typically varying in terms of phonology, semantics, speech fluency and, when assessed, non-language cognitive skills (Butler et al., 2014, Halai et al., 2017, Halai et al., 2018a, Schumacher et al., 2019). In this formulation the classical aphasia labels (e.g., conduction, Broca's, etc.) do not represent categories per se, but rather are verbal pointers to a subregion in the multidimensional space. By way of analogy, one can think of patients as colour hues across the red, green and blue (RGB) colour space. It is possible to recognise clear differences (such as yellow (e.g., Broca) vs. blue (Wernicke), etc.) but also to capture the graded and unbounded variations between colours (e.g., there are many types of blue, its boundary with greens or violets is unclear, there are many hues (e.g., teal, maroon, etc.) that are hard to classify uniquely, and perceivers (cf. clinicians/researchers) have slightly different definitions for each colour (cf. clinical label)).

Accordingly, some key aims of the current study were: (a) to test if the same approach can be applied to PPA (in contrast to other studies where methods capable of capturing graded variation have only been used as an intermediate step towards categorising proposed subtypes of PPA (Hoffman et al., 2017, Mesulam et al., 2009)); (b) to compare the multidimensional spaces for PSA and PPA; (c) to test if a single multidimensional space can
be formed for PSA and PPA to allow direct, intragroup and intergroup comparisons. Importantly, the aim of situating PSA and PPA in a shared multidimensional space was not to differentiate between these aetiologies. Rather, we used this shared space as a platform for a larger-scale direct comparison without selecting specific subtypes or cognitive/linguistic processes (as mentioned above). These aims were tackled through two large PSA and PPA cohorts (inclusive of typical and mixed cases), both completing large-scale, detailed neuropsychological and aphasiological test batteries.

**Methods**

We initially applied principal component analysis (PCA) to PPA and PSA separately. This allowed us to compare qualitatively the resultant multidimensional space for each patient group without forcing the two groups into a single space. Given that the two group-specific PCA results were very similar in form, available PSA patients were re-assessed using a shared test battery derived from the PPA test battery, so that all patients could be entered simultaneously into a unified PCA. This enabled direct comparisons of both intergroup and intragroup variations.

**Patients**

All patients were recruited non-selectively (with respect to subtype-level behavioural presentation) to sample the full space and severities of behavioural impairments in both PPA and PSA. Although diagnostic subtype labels were applied for descriptive purposes, the inability to apply a single diagnostic label was not grounds for exclusion in either cohort. Demographic details are shown in Table 2.1.

Seventy-six people with chronic PSA were prospectively recruited from community groups and speech and language therapy services in the North West of England. Patients were included if they reported a single left hemisphere stroke at least 12 months prior to assessment and were native English speakers. A portion of the PSA cases have been reported by Butler et al. (2014) and Halai et al. (2017) (31/70), and by Halai et al. (2018b) (70/76), Woollams et al. (2018) and Schumacher et al. (2019). All patients were classified into diagnostic subtypes by application of the Boston Diagnostic Aphasia Examination (Kaplan, 1983). All patients provided informed consent under approval from the North West Multi-Centre Research Ethics Committee, UK. Thirty-four of the 76 PSA cases were available for re-testing on the shared battery for the unified PCA on PSA and PPA. Of these 34 cases, 15 had anomic aphasia, 5 had Broca’s aphasia, 3 had conduction aphasia, 5 had global aphasia, 5 were classified as mixed non-fluent aphasia, and 1 had transcortical motor aphasia.

Forty-six people with PPA were prospectively recruited from memory clinics at Addenbrooke’s Hospital, University of Cambridge (UK), as part of a longitudinal study of PPA (Sajjadi, 2013). These cases have previously been reported by Sajjadi et al. (2012b, 2012c, 2012a, 2014) and Hoffman et al. (2017). Patients with PPA were recruited on the
basis of meeting the core criteria for PPA (Mesulam, 2001a) then classified into a diagnostic subtype by application of the Gorno-Tempini et al. (2011) criteria or given the label ‘mixed PPA’ if unclassifiable. Exclusion criteria included other causes of aphasia (e.g., non-neurodegenerative pathology), non-native English speakers and any other neurological or major psychiatric illness. The PPA dataset comprised data from two longitudinal rounds of testing to assess change over time. On average, the second round of data was collected after 12.7 months (standard deviation: 0.9 months). Following Lambon Ralph et al. (2003), the participants who had scores for both rounds were treated as pseudo-independent observations. This resulted in a total of 82 data points for analysis. Of these 82 observations, 26 were mPPA, 24 were PNFA, 28 were SD and 4 were logopenic PPA. This approach was employed because PCA is a data-hungry method (Guadagnoli and Velicer, 1988) which benefits from having adequate sampling of as much of the potential PPA ‘space’ as possible. All patients, or next of kin where appropriate, provided informed consent under approval from the Cambridge Regional Ethics Committee.

Neuropsychological assessments

Post-stroke aphasia test battery

The tests included in the PSA test battery are shown in Supplementary Figure 2.4, and described in Halai et al. (2017) and Butler et al. (2014). Briefly, the battery assessed connected speech, comprehension of grammar, auditory discrimination, repetition, semantic knowledge, naming, working memory and attention/executive function.

Primary progressive aphasia test battery

The tests included in the PPA test battery are shown in Supplementary Figure 2.5, and described by Sajjadi et al. (2013). Briefly, this battery assessed connected speech, comprehension of grammar, grammatical ability in sentence production, repetition, semantic knowledge, naming, phonological discrimination, working memory, attention and executive function, visuospatial skills, and oro-buccal and limb praxis.

Shared battery

In order to establish the shared multidimensional space of PSA and PPA, available PSA cases were re-tested on a shared test battery which was derived from the PPA test battery. Thirty-three tests were including in the shared battery, shown in Supplementary Figure 2.6. This battery assessed attention and executive function, repetition, sentence comprehension and production, semantic memory, visuospatial skills, praxis, connected speech, naming, and phonological discrimination.

Data analysis

All raw behavioural scores were converted to percentages. For measures without a fixed maximum score, scores were converted to a percentage of the maximum score across the relevant cohort or both cohorts for the unified PCA. Missing data were imputed using probabilistic principal component analysis (PPCA) (Ilin and Raiko, 2010). This approach was
chosen as the results were stable when compared to versions of the analyses without imputation (i.e., list-wise exclusion analysis). PPCA requires that the number of components to be extracted is specified a priori, so a k-fold cross validation approach (Ballabio, 2015) was used to choose the number of components giving the lowest root mean squared error for held-out cases over 1000 permutations. This approach was also used to select the optimal number of components for subsequent PCA using the imputed dataset. The imputed datasets were entered into PCAs (conducted in SPSS 23), with varimax rotation to aid cognitive interpretation of the extracted dimensions. This interpretation was based on the core aspects of the tests with the largest loadings onto each factor. The factor labels necessarily capture less information than the test loadings (which are provided in Supplementary Materials) but serve as a useful short-hand. The adequacy of the sample size for each PCA was determined using Kaiser Meyer Olkin measure of sampling adequacy.

The separate PCAs for PPA and PSA could not be compared directly since they did not share the same tasks, so they were compared qualitatively by analysing the type of tasks that loaded most heavily onto each extracted component. This approach was also used to compare the separate PCAs to the unified PCA. However, since the unified PCA was conducted on data from both groups on the shared battery, this made direct, quantitative, intra- and intergroup comparisons possible. To put the relative regions of the multidimensional space occupied by PSA and PPA into perspective, control norms were projected into the unified PCA space by normalising to the patient mean and standard deviation, and then using the factor coefficients to generate factor scores for an average control participant.

Finally, formal analyses were conducted to test for the presence of subgroup categories (i.e., subgroups with relatively high intra-group homogeneity and distinct intergroup differences). The motivation for this analysis was as follows: working under the hypothesis that the structure of variation in PSA and in PPA is driven by graded variation along multiple dimensions means that methods like cluster analysis (e.g., k-means clustering) would be inappropriate for detecting potential graded variation. Yet, visual inspection of the scatter plots defined by the extracted PCA dimensions (e.g., Figure 2.2) showed many regions of extensive overlap but also (for SD) some more uniquely occupied regions. Therefore, we sought to quantify this by conducting a form of data-driven classification analysis within the graded multidimensional space, rather than using formal cluster analysis. We note that the principal dimensions were not intended as a new way to categorise patients. Instead, we took it to be the case that if one or more subgroups formed a true category, they would be represented in the PCA multidimensional space as a homogenous group of data points, and it would be possible to define formal diagnostic boundaries with the other subgroups in terms of cut-off scores on each extracted dimension.
To investigate this, formally, the unified PCA was systematically swept to find the combination of cut-off values across all dimensions that gave the highest sensitivity index (d prime – d’) value per diagnostic subtype. Crucially, the calculations of sensitivity were only within aetiology, i.e., only considering subtypes from the same cohort; the aim of this analysis was not to differentiate PSA and PPA. PCA solutions are always scaled using z-scoring, and in this study the dimensions ranged from approximately -3 to +2 with zero representing the centre of the patient cohort. These dimensions were swept iteratively at intervals of 0.05. The d’ equation was adapted to account for extreme values (0 or 1) for the rate of hits or false alarms (Macmillan and Kaplan, 1985), resulting in the maximum d’ value achievable being 4.65; thus a d’ value near 4.65 would be suggestive of distinct categorical boundaries and within-group homogeneity. To establish the likelihood of achieving these d’ values by chance, diagnostic group-membership was randomised within aetiology and d’ recalculated over 10,000 iterations to give a distribution of d’ values.

The combination of cut-off values along each dimension that gave the maximum d’ value (i.e., highest possible sensitivity) were then treated as diagnostic ‘criteria’ for new data-driven diagnostic groups. The hits from the d’ analysis represent cases whose factor scores correctly met the cut-off values for their own data-driven diagnostic group. The false alarms represent cases whose factor scores incorrectly met the cut-off values for any other data-driven diagnostic group, i.e., cases who were misclassified. These misclassifications occurred despite the cut-offs representing the best possible (highest sensitivity) between-group boundaries that could be found in the iterative sweep through the entire multidimensional space. The distribution of misclassifications amongst subtypes of each aetiology was extracted from the false alarms associated with each data-driven diagnostic group. It is important to note that it was possible for a single case’s factor scores to meet the cut-off values for more than one data-driven diagnostic group simultaneously (or none), and this may or may not have included their own group.
Results

Demographics

The demographic details for all participants are displayed in Table 2.1.

Table 2.1 – Demographic details per subtype of the post-stroke aphasia and primary progressive aphasia cohorts.

<table>
<thead>
<tr>
<th>Group</th>
<th>Subtype</th>
<th>N (F)</th>
<th>Age (years)</th>
<th>Education (years)</th>
<th>Time with aphasia (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA</td>
<td>Anomia</td>
<td>30 (11)</td>
<td>63.8 (13.4)</td>
<td>12.4 (2.7)</td>
<td>4.2 (4.3)</td>
</tr>
<tr>
<td></td>
<td>Broca</td>
<td>13 (1)</td>
<td>62.7 (13.0)</td>
<td>11.9 (1.8)</td>
<td>4.3 (3.4)</td>
</tr>
<tr>
<td></td>
<td>Conduction</td>
<td>4 (1)</td>
<td>62.0 (10.7)</td>
<td>13.8 (3.2)</td>
<td>1.8 (0.9)</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>9 (0)</td>
<td>66.4 (9.0)</td>
<td>11.3 (0.7)</td>
<td>5.9 (4.6)</td>
</tr>
<tr>
<td></td>
<td>Mixed non-fluent</td>
<td>16 (4)</td>
<td>68.1 (9.0)</td>
<td>11.5 (1.0)</td>
<td>6.3 (4.9)</td>
</tr>
<tr>
<td></td>
<td>TMA</td>
<td>2 (1)</td>
<td>74.5 (2.1)</td>
<td>11.0 (0.0)</td>
<td>6.8 (4.1)</td>
</tr>
<tr>
<td></td>
<td>TSA</td>
<td>1 (0)</td>
<td>63.0</td>
<td>12.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Wernicke/conduction</td>
<td>1 (1)</td>
<td>77.0</td>
<td>16.0</td>
<td>2.8</td>
</tr>
<tr>
<td>PPA</td>
<td>Logopenic</td>
<td>2 (1)</td>
<td>71.0 (4.2)</td>
<td>11.0 (2.8)</td>
<td>2.0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Mixed PPA</td>
<td>16 (12)</td>
<td>72.7 (5.2)</td>
<td>10.8 (1.9)</td>
<td>3.3 (1.4)</td>
</tr>
<tr>
<td></td>
<td>PNFA</td>
<td>12 (7)</td>
<td>69.3 (7.3)</td>
<td>13.0 (3.8)</td>
<td>3.2 (1.4)</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>16 (8)</td>
<td>67.1 (8.5)</td>
<td>13.9 (3.3)</td>
<td>4.2 (1.3)</td>
</tr>
</tbody>
</table>

Data presented as mean (standard deviation). Abbreviations: PSA – post-stroke aphasia; PPA – primary progressive aphasia; TMA – transcortical motor aphasia; TSA = transcortical sensory aphasia; PNFA – progressive non-fluent aphasia; SD – semantic dementia.

Principal component analysis

Post-stroke aphasia

The PCA for the PSA cohort was robust (Kaiser Meyer-Olkin = 0.84) and produced a 4-factor rotated solution which accounted for 76.7% of variance in PSA patients’ performance.
Measures loading heavily onto the first factor were tests of repetition (PALPA words, non-words), naming (Boston, Cambridge), phonological working memory (digit span), auditory comprehension (CAT sentence comprehension), and phonological sensitivity (PALPA minimal pairs). These tests all require phonological processing; hence we called this factor ‘Phonology’. The strong loadings from the naming tests onto this factor are likely to be driven by the fact that many of the cases in this PSA cohort have a core phonological processing impairment (hence this Phonology factor explained the greatest amount of variance in the PCA) and so the phonological aspect of naming is compromised in these patients.

The second factor had strong loadings from the two measures designed to assess attention and executive function (Brixton, Raven’s). Other tests not designed to measure executive function per se also had strong loadings onto this factor (e.g., minimal pairs, spoken word-to-picture matching, etc.). This probably reflects the fact that tasks designed to assess various language activities also call upon generalized attention and executive skills (e.g., to compare verbal stimuli, decide between responses, etc.). This is true for the semantic tests (aligning with the fact that semantic cognition requires access to semantic representation but also executively-related processes (Jefferies and Lambon Ralph, 2006, Thompson et al., 2018) and also with respect to the working memory, abstract reasoning and problem-solving requirements in other language tests (e.g., sentence comprehension and minimal pairs). The nature of PCA means that it decomposes and orthogonalizes these sources of variation. As a result, individual tests can have strong loadings across multiple extracted factors and each factor points towards a shared underpinning process. Thus, for this second factor, whilst spanning different aspects of language and cognition, these tests all share the feature of requiring attentional or executive processing skills; hence, we called this factor ‘Executive Function’.

Measures with strong loadings onto the third factor included speech rate (words per minute) and speech quanta (total number of words). We called this factor ‘Speech Production’. We note that the Camel & Cactus test, an executively-demanding test of semantic associative relationships, also had a very strong loading onto this factor. This result is surprising and has not occurred in our previous investigations, where it has loaded onto the executive and semantic factors (Butler et al., 2014, Halai et al., 2017). We could speculate that this result might reflect variation in another form of executive process (distinct from the executive process that seems to be captured by the second factor), which might be involved in (i) iteratively generating and assessing semantic associations (Camel & Cactus and synonym judgement), (ii) generating speech (words per minute and total words produced), (iii) generating and monitoring ‘chunks’ to complete the backwards digit span task. Indeed, Schumacher et al. (2019) found that the Camel & Cactus test loaded onto an ‘Inhibit-Generate’ executive component. However, without more measures of attention and
executive function it is not possible to test these speculations. We chose to give this third
factor the subjective label of ‘Speech production’ given that the strongest loadings are from
words per minute and total number of words.

Measures with strong loadings onto the fourth factor were tests of semantic knowledge
(synonym judgment, word-to-picture matching) and semantic richness of speech (mean
length per utterance). Furthermore, measures of naming (Boston, Cambridge) and sentence
comprehension (CAT) also had moderate factor loadings onto this factor. These tests all
require semantic knowledge; hence we called this factor ‘Semantics’.

Primary progressive aphasia

The PCA for the PPA cohort also generated a robust result (Kaiser Meyer-Olkin=0.85) with a
5-factor rotated solution which accounted for 72.4% of variance (F1 = 23.7%, F2 = 17.8%,
F3 = 14.5%, F4 = 9.8%, F5 = 6.7%). The factor loadings of each behavioural assessment
onto the extracted components are shown in Supplementary Figure 2.5.

The tests loading onto the first factor all required retaining phonological information (e.g.,
single digits, words, numbers, whole sentences) in mind; hence we called this factor
‘Phonological Working Memory’. These measures included tests of phonological sensitivity
(non-word minimal pairs), attention and executive function (digit span forwards and
backwards, letter span similar and dissimilar phonemes), repetition (words, non-words and
sentences), sentence comprehension (SECT, TROG) and cube counting (VOSP).

The second factor comprised heavy loadings from tests relying on semantic knowledge;
hence we called this factor ‘Semantics’. These tasks included tests of semantic knowledge
(Cambridge naming, Point from Repeat and Point, Category Fluency), semantic association
(Camel & Cactus), recognition of irregular words, and sentence comprehension (SECTV).

The third factor was characterised by strong loadings from measures of speech rate (words
per minute) and speech quanta (total number of words), in addition to oro-buccal praxis. A
test of executive function requiring drawing and counting (TMT-A) also had high loadings
(note, patients often count under their breath or out loud to complete the TMT-A).
Accordingly, this factor appeared to capture the motor aspect of speech, hence we called
this factor ‘Motor Speech Production’.

Visuospatial tests of executive function loaded heavily onto the fourth factor. Specifically,
tests of switching (TMTB), counting and visual imagery (VOSP), and copying and
visuospatial recall (Rey Complex Figure) had high loadings on this factor, hence we called
this factor ‘Visuo-Executive Function’.

Loadings onto the fifth factor were dominated by tests of sentence production (MAST),
measures of semantic richness of speech (mean length per utterance) and generation of
items (letter fluency from the Addenbrooke’s Cognitive Examination – Revised (Mioshi et al.,
2006)). Having accounted for motor speech production, semantics and executive demands
in earlier factors, the remaining aspect of these tests which might be captured in this final independent factor could be the generative aspect of speech production. Hence, we called this factor ‘Speech Generation’.

*Unified PCA on the shared battery*

Given that the two group-specific batteries and PCAs generated similar types of dimensions (phonology, semantics, executive skill and aspects of speech production), a formal direct comparison through a shared battery and single PCA spanning both groups was both merited (i.e., there was *prima facia* evidence of shared symptoms and variations) and permitted formal intra-group and intergroup comparisons by enabling inclusion of all patients across a single multidimensional space. The unified PCA was again robust (Kaiser Meyer-Olkin=0.88), with a 4-factor rotated solution accounting for 67.4% of variance in patients’ performance (F1 = 23.5%, F2 = 16.6%, F3 = 14.8%, F4 = 12.6%), and bore a strong relationship with the factors identified in the group-specific test batteries. The factor loadings of each behavioural assessment onto the extracted components are shown in Supplementary Figure 2.6.

The first factor had high loadings from tests of repetition (words, non-words, sentences), phonological sensitivity and attention (digit and letter spans, non-word minimal pairs), and sentence comprehension (auditory and visual). These tests all require intact phonological processing, hence we called this factor ‘Phonology’.

There were high loadings on the second factor for tests of semantic knowledge (Cambridge naming, pointing from the Repeat and Point test), generation of items in a semantic category (category fluency from the ACE), sentence comprehension (SECT) and address recall and recognition from the ACE. Hence, we called this factor ‘Semantics’.

Tests of attention and executive function in the visuospatial domain (VOSP, TMT, Rey Complex Figure) all loaded heavily onto the third factor. As above, we called this factor ‘Visuo-Executive Function’.

The fourth factor had high loadings from measures of speech quantity (words per minute, total number of words and mean length per utterance). Measures of praxis (Oro-buccal and limb) and phonological working memory (digit span backwards) also had high loadings onto this factor. Given that phonological ability and executive functions have been accounted for already, this factor probably captured the speech production element of the digit span test (patients often repeat the string of digits to themselves before reporting them backwards). These tests therefore all require production of speech, and coupled with the loadings from the praxis tests, we interpreted this as a ‘Motor Speech Production’ factor.

The Visuo-Executive Function and Motor Speech Production factors had strong negative loadings from the TMTB and TMTA response times, respectively. We re-ran the analysis without these measures to ensure the result was stable and that the negative loadings were
not an artefact of a coincidental correlation with general motor abilities. Pearson correlations between the original factors and their corresponding updated factors following the removal of the TMT response time measures were very high (F1 vs F1: 0.999, F2 vs F2: 0.997, F3 vs F3: 0.996, F4 vs F4: 0.991), showing that the PCA result was unchanged.
Intergroup comparisons

To illustrate the components extracted in each PCA, exemplar tests with strong and relatively unique loadings onto each factor across the three PCAs are plotted together in Figure 2.1 (the full plots of all tests loadings on all PCA factors are shown in the Supplementary Figures). The specific example test chosen differed across PCAs due to the different test batteries, but where possible the same or a similar measure was chosen. This figure highlights the similarity of the components extracted for both forms of aphasia, whether separately or in combination.

Figure 2.1 – Intergroup comparison of the underlying dimensions of variance in PSA and PPA.

Bars represent the factor loadings of exemplar tests onto each extracted factor. Factor loadings represent the weighting of each test on each factor and were used to suggest cognitive interpretations of the factors. The patterns of the bars represent the different PCAs; the PPA PCA extracted two speech production components, which are shown in different patterns on the Motor Speech Production panel. Abbreviations: PSA – post-stroke aphasia; PPA – primary progressive aphasia; PCA – principal component analysis; VOSP – Visual Object and Space Perception battery (Warrington and James, 1991); MAST – Make a Sentence Test (Billette et al., 2015).
Direct intra- and intergroup comparisons were possible in the shared multidimensional space of the unified PCA. Figure 2.2 plots the patients and their aphasia classifications (PSA in blue and PPA in red markers) into the four-dimensional factor space (Panel A maps the phonology and semantics factors, Panel B speech production vs. visuo-executive skill factors). Four key observations can be gleaned from these scatterplots: (i) intra-group graded differences: for both PPA (except semantic dementia, see below) and PSA there is considerable variation across cases within each subtype of aphasia and also overlap between the groups (e.g., conduction and anomic aphasia or PNFA and mixed PPA); (ii) intergroup differences: with regards to semantics and phonology the PSA and PPA cases are fully overlapping reflecting the clinical observation that the two aetiologies share many language symptoms; (iii) the two aetiologies are strongly separated in terms of speech fluency and co-occurring visuo-executive skills with the PSA cases dominating the space denoting poorer fluency yet better visuo-executive skills (lower right quadrant in Panel B). All forms of PSA (even those referred to as “fluent”) were less fluent that the PPA patients (with the exception of the most severe PNFA and mixed cases), whilst only the SD subset were able to match the PSA on visuo-executive skills; (iv) by eye, the only group which might form a coherent and separated cluster (cf. a true category) are those with SD (yellow crosses) in that they appear to uniquely occupy the combination of moderate-to-severe semantic impairment with good phonology (i.e., lower right quadrant in Panel A) and good visuo-executive function and speech fluency (i.e., upper right quadrant in panel B). We tested formally whether SD and any other groups form a true category in the subsequent analysis.
Figure 2.2 - Regions of the shared multidimensional space of PSA and PPA occupied by each diagnostic subtype.

Factor scores of all patients were plotted along all pairs of components extracted from the unified PCA. The origin is the mean of all patients. The factor scores are an expression of how many standard deviations a patient’s performance is from the group mean. The region of space reflecting preserved performance was calculated by projecting control norms into the patient space and is shaded in grey. PSA subtypes are blue-spectrum colours, PPA are red-spectrum colours. Abbreviations: PSA - post-stroke aphasia; PPA - primary progressive aphasia; TMA - transcortical motor aphasia; SD - semantic dementia; PNFA - progressive non-fluent aphasia.
**Intragroup graded variation**

For each subtype within each aetiology, the best combination of ‘diagnostic’ values across all four dimensions was derived using a data-driven search (see Methods; all values are displayed in Supplementary Table 2.3). These values were treated as cut-offs defining new data-driven diagnostic groups, which were labelled according to the subtype from which the cut-offs were derived. An illustration of the data-driven diagnostic cut-offs for SD is shown in Figure 2.3.

The pattern of hits and misclassifications associated with each combination of diagnostic cut-offs for PPA subtypes is shown in Table 2.2. The pattern of hits and misclassifications for PSA subtypes is shown in Supplementary Table 2.4, as these results will need validating in a larger cohort; in order to include a heterogeneous cohort reflecting the true phenotypic space of PSA, subtypes of PSA were included in this study even if they comprised only a single case. However, in terms of assessing whether the subtypes of PSA meet the assumptions of a true category, larger sample sizes will be needed to fully answer this question.

The data presented in Table 2.2 are the percentages of patients, from each clinical diagnostic subgroup of PPA, whose factor scores met the data-derived cut-off values for each diagnostic group. The rows represent the ‘real’ clinical diagnostic categories for patients in this study. The ‘Hits’ column represents the percentage of patients meeting the data-driven cut-off values for their own diagnostic group. The columns under ‘Misclassifications’ represent the percentage of cases whose factor scores (a) met the cut-off values for a different (i.e., incorrect) diagnostic group; (b) did not meet the cut-off values for any of the possible data-driven diagnostic groups; (c) met the cut-off values for more than one data-driven diagnostic group (e.g., their own group and one other group). These ‘Misclassifications’ columns are not mutually exclusive and thus cases falling into more than one classification are tabulated in the ‘>1’ column; consequently, the row totals do not add up to 100%.

For example, the optimum cut-off values for SD were highly selective for SD, with 100% of the SD cases’ factor scores meeting these values (‘Hits’ column). Furthermore, as can been seen from the ‘Misclassifications – SD’ column, there were no misclassifications of patients from other diagnostic groups as SD. This corresponds to the highest d’ value of 4.46 (p < .001) for SD and suggests that the SD cases from which the data-driven diagnostic criteria were derived show within-group homogeneity and clearly distinct between-group boundaries. This corroborates our earlier qualitative interpretation of SD as occupying a unique area in the multidimensional space from the unified PCA. Due to the data-driven criteria for other PPA subtypes being less selective for their target subtype, some SD cases were misclassified (columns ‘Misclassifications – mPPA’ and ‘Misclassifications – PNFA’); accordingly, in the ‘Misclassifications - >1’ column, 28.6% of the SD cases met both the SD cut-offs and the cut-offs for either mPPA or PNFA.
Figure 2.3 - Data-driven diagnostic cut-off values for semantic dementia.

Using a data-driven sweep at intervals of 0.05 through the entire four-dimensional space, the combination of cut-off values giving optimum sensitivity for semantic dementia was isolated. Applying the simultaneous combination of these four-dimensional cut-off values as diagnostic criteria (dashed lines) gave perfect selectivity for semantic dementia. This implies that semantic dementia shows within-group homogeneity and distinct between-group differences, suggestive of a true diagnostic category. This process was repeated for all subtypes of PPA and PSA within each aetiology (cut-off values and d' values per subtype in Supplementary Table 2.3). Abbreviations: SD – semantic dementia; PNFA – progressive non-fluent aphasia.
The data-driven diagnostic criteria for PNFA ($d' = 2.03, p < .001$) were much less selective as not only did they incur misclassifications of all other subtypes, but they also failed to classify all the PNFA cases correctly (hit rates around 90%). This shows that even the optimal data-driven diagnostic criteria for PNFA were insufficiently selective, implying that PNFA cases do not display within-group homogeneity and distinct between group boundaries like SD.

The classification results for lvPPA and mPPA are presented here for completeness, with the caveats that (a) the lvPPA result was derived from a very small sample size and so will need validating with a much larger cohort, and (b) mPPA does not represent an actual subtype category of PPA (instead it represents the label given to cases who do not meet the criteria for any proposed category). This means that the data-driven diagnostic criteria for mPPA would not be expected to have any selectivity for this inherently heterogeneous group. The lvPPA data-driven diagnostic criteria ($d' = 3.38, p < .001$) had a perfect hit rate but some misclassifications of PNFA cases, resulting in a lower $d'$. Consistent with the nature of cases given the mPPA subtype label, the data-driven diagnostic criteria for this group ($d' = 2.10, p < .001$) also showed low selectivity. The mPPA criteria failed to capture all the mPPA cases and incorrectly captured cases from all other subtypes.
Table 2.2 – Distribution of misclassifications between clinical and data-driven diagnostic PPA groups.

<table>
<thead>
<tr>
<th>Clinical diagnostic groups (N)</th>
<th>Hits</th>
<th>Data-driven diagnostic groups</th>
<th>Misclassifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>lvPPA (4)</td>
<td>100.0</td>
<td>IvPPA</td>
<td>100.0</td>
</tr>
<tr>
<td>mPPA (26)</td>
<td>92.3</td>
<td>mPPA</td>
<td>38.5</td>
</tr>
<tr>
<td>PNFA (24)</td>
<td>91.7</td>
<td>PNFA</td>
<td>0.0</td>
</tr>
<tr>
<td>SD (28)</td>
<td>100.0</td>
<td>SD</td>
<td>3.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>lVPPA</th>
<th>mPPA</th>
<th>PNFA</th>
<th>SD</th>
<th>None</th>
<th>&gt;1</th>
</tr>
</thead>
<tbody>
<tr>
<td>lvPPA (4)</td>
<td>100.0</td>
<td>/</td>
<td>50.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>mPPA (26)</td>
<td>92.3</td>
<td>0.0</td>
<td>/</td>
<td>38.5</td>
<td>0.0</td>
<td>3.8</td>
</tr>
<tr>
<td>PNFA (24)</td>
<td>91.7</td>
<td>4.2</td>
<td>20.8</td>
<td>/</td>
<td>0.0</td>
<td>4.2</td>
</tr>
<tr>
<td>SD (28)</td>
<td>100.0</td>
<td>0.0</td>
<td>25.0</td>
<td>3.6</td>
<td>/</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Data expressed as percentages of the total number of cases per clinical diagnostic group. Abbreviations: lvPPA – logopenic primary progressive aphasia; mPPA – mixed primary progressive aphasia; PNFA – progressive non-fluent aphasia; SD – semantic dementia.

The cut-off values giving optimum sensitivity for each diagnostic group were treated as data-driven diagnostic criteria. Rows represent ‘real’ clinical diagnostic categories. The ‘Hits’ column represents the percentage of patients meeting the data-driven cut-off values for their own data-driven diagnostic group. The columns under ‘Misclassifications’ represent the percentage of cases whose factor scores (a) met the cut-off values for a different data-driven diagnostic group; (b) did not meet the cut-off values for any of the data-driven diagnostic groups; (c) met the cut-off values for more than one data-driven diagnostic group. These ‘Misclassifications’ columns are not mutually exclusive, so row totals do not add up to 100%.
Discussion

This study had two principal aims: (i) to undertake a large-scale direct comparison of PPA and PSA utilising detailed aphasiological and neuropsychological test batteries, and (ii) to reconsider the phenotype differences across patients with PSA and PPA in terms of graded variations along multiple principal dimensions.

The results confirm that there is meaningful, coherent structure in the language-cognitive variations across PPA and PSA patients. Rather than conceptualising such variations as mutually-exclusive categories, the results indicate that the patients’ variations reflect multiple, continuous, graded dimensions. This alternative approach has multiple advantages: (a) it is able to capture the patterns of overlap between different ‘subtypes’ of PPA and PSA (e.g., overlap in phonological impairments in many PPA and PSA cases) as well as their clear differences; (b) it captures the variations in performance within each ‘type’ of PPA and PSA; and (c) it meaningfully situates the ‘mixed’ aphasic patients alongside the other cases to generate a complete clinical picture of PPA and PSA. This is important given that there are high numbers of ‘mixed’ cases in everyday clinical practice. It is, perhaps, important to note that these dimensions are not new categories but rather each patient represents a specific point in the graded multidimensional space.

The cognitive and language impairments in PSA and PPA, both when considered in isolation and when considered in a single unified framework, could be captured by four main dimensions of underlying variation: phonology, semantics, speech production/motor output fluency, and executive-cognitive skill. This finding is a direct replication of what has been found previously for this PSA cohort (Butler et al., 2014, Halai et al., 2017) and by numerous international groups (Kümmerer et al., 2013, Lacey et al., 2017, Mirman et al., 2015b, Mirman et al., 2015a, Tochadse et al., 2018), and found to be statistically stable across different sample sizes and assessment batteries (Halai et al., 2018a). These studies have used lesion-symptom mapping methods to show that the principal components are associated with neural correlates that support the labels applied (e.g., components labelled ‘Phonology’ having neural correlates in left posterior perisylvian cortical (e.g., superior temporal gyrus) and subcortical regions (e.g., arcuate fasciculus (Butler et al., 2014, Halai et al., 2017), and dorsal parietal white matter (Lacey et al., 2017)), which have previously been shown to be involved in phonological processing). Although outside the scope of this study, neuroimaging information could help to elucidate and further delineate the underlying principal dimensions of variance in both forms of aphasia. Aphasia can be caused by brain injury to cortical but also subcortical brain areas (Naeser et al., 1989, Naeser et al., 1987, Hillis et al., 2002, Hillis et al., 2004). Therefore, the location of the lesion/atrophy is a critical piece of information to help understand the mapping between behavioural dimensions and neural substrates (Naeser and Palumbo, 1994). The location of the lesion/atrophy also relates to functional connectivity changes associated with aphasia (Yang et al., 2016, Ranasinghe et al., 2017). Future research could combine the PCA framework employed in
this study with single- or multi-modality imaging information (such as white matter integrity or functional connectivity) to explore how the underlying nature of the brain injury (i.e., relatively discrete lesion vs. relatively diffuse atrophy) leads to similar/differing neural changes and consequently to similar/differing behavioural symptoms in PSA and PPA. Another avenue for future research could be to include longitudinal neuropsychological data in this PCA framework, in order to contrast the temporal profiles of recovery (in a less chronic PSA cohort) vs. degeneration. This would inform our understanding of how different aetiologies of brain damage result in a changing aphasic profile in these populations.

The fact that the same underlying dimensions were found for PPA as well as PSA indicates that these dimensions might reflect core “primary systems” for language activities (Patterson and Lambon Ralph, 1999, Ueno et al., 2014, Woollams et al., 2018). Past work has associated these primary systems with different brain areas: phonological processing and working memory with posterior superior temporal lobe and supramarginal gyrus (Paulesu et al., 1993); semantic representation with anterior temporal lobe (ATL) (Patterson et al., 2007, Lambon Ralph et al., 2017); speech programming and fluency with premotor cortex and key underpinning white matter pathways (Basilakos et al., 2014); and executive functions with frontoparietal networks (Jurado and Rosselli, 2007, Marek and Dosenbach, 2018). As these regions can be affected in both middle cerebral artery PSA (Phan et al., 2005) and PPA (Gorno-Tempini et al., 2004), the similarity of their phenotypic spectra could reflect varying degrees of impairment to these core primary systems.

Plotting all patients’ factor scores into the shared multidimensional space showed that the non-SD PPA and PSA cases occupied an almost completely overlapping region of the Phonology-Semantics space. This contrasts with the SD cases who occupied an exclusive region of the multidimensional space, signifying their selective semantic impairment in the context of relatively preserved phonological abilities, coupled with motor speech production and executive function that are comparable to healthy controls. This might reflect the fact that SD arises from atrophy in extra-sylvian, ATL regions (Snowden et al., 1989, Hodges et al., 1992, Rosen et al., 2002), whereas the other forms of PPA and PSA are associated with damage to perisylvian cortical and subcortical regions (Grossman and Irwin, 2018, Hillis et al., 2002, Hillis et al., 2004).

In the space corresponding to Visuo-Executive Function vs. Motor Speech Production, there was separation of the two aetiologies. PSA patients occupied the region signifying less fluent speech production combined with relatively unimpaired visuo-executive ability. Most PPA patients showed the reverse pattern, although some PNFA and mPPA cases showed the combination of poor fluency and poor executive function. This result agrees with previous direct comparisons restricted to the non-fluent subtypes of PSA and PPA (Patterson et al., 2006a). This separation is clinically interesting and important as it indicates that certain symptom terms – e.g., fluency – are not used in the same way across patient types; thus, many non-fluent progressive aphasics were more fluent than the “fluent” PSA cases (e.g.,
anomic and conduction aphasics). This may be relevant for clinical professionals who work with people with PSA and with people with PPA; if assessments/tools at their disposal are targeted towards 'non-fluent' aphasias then it may be useful to have a formal understanding of how the term ‘fluency’ is applied across PSA and PPA.

The separation in terms of Visuo-Executive Function could reflect an aetiology-driven difference in the neural substrates vulnerable to damage in stroke vs. neurodegeneration; cognitive functions supported by regions at the edges of/outside the territory of the middle cerebral artery (MCA) would be less likely to be impaired in PSA than perhaps in some forms of PPA. For example, the multi-demand frontoparietal executive system (Marek and Dosenbach, 2018), and posterior cingulate and other medial regions which support executive function and attention (Jurado and Rosselli, 2007) are situated at the edges/outside of the MCA-perfused regions (Phan et al., 2005), potentially leading to relatively spared Visuo-Executive Function in our PSA cohort.

In addition to facilitating intergroup comparisons, the PCA method revealed graded intra-group differences. The subtypes of non-SD PPA and PSA occupied only partially differentiated positions within the four-dimensional space, with considerable variation within each “subtype” and overlap of cases across subtypes. This could reflect the overlapping atrophy/lesions in and around the cortical and subcortical perisylvian language regions in these forms of aphasia. Again, these findings indicate that phenotypic variations in non-SD PPA and PSA are unlikely to reflect different categories but rather graded variations along these dimensions. These graded differences can only be accounted for in categorical classification systems by using ‘mixed’ classifications (Wertz et al., 1984, Wicklund et al., 2014, Sajjadi et al., 2012a), but the methods in the current study were able to account for graded variation in a single multidimensional framework comprising four, clinically-intuitive underlying dimensions.

Based on this framework, the current diagnostic subtype labels can be reconceptualised as pointers towards particular regions of the multidimensional space, rather than labels for mutually-exclusive clinical categories. This approach does not preclude the fact that some labels might be pointers for more exclusive regions of space (e.g., SD or global PSA) than others (e.g., anomia or PNFA).

In fact, the concept of ‘SD’ seems to be a uniquely useful pointer for the exclusive region of the multidimensional space occupied by these cases. This aligns with (i) the original descriptions of SD, in particular the selective nature of their semantic impairment (Warrington, 1975, Hodges et al., 1992, Snowden et al., 1989), and (ii) previous work showing that SD is distinct from other forms of PPA; Bisenius et al. (2017) found that SD was the most readily differentiable subtype of PPA using Support Vector Machine approaches to evaluate the consensus criteria for PPA. Hoffman et al. (2017) applied k-means clustering to behavioural data in PPA and found that of their three-cluster solution,
only one cluster was selective for a particular subtype of PPA and this was the SD cluster. Furthermore, by plotting PSA and PPA in the same space we provide support for previous work showing that semantic impairments in SD are unlike those found in PSA (Jefferies and Lambon Ralph, 2006, Lambon Ralph et al., 2017). This result probably reflects the fact that the distribution of damage in SD is distinctly different from those in non-SD PPA and PSA phenotypes. SD cases have hypometabolism and atrophy centred on the ATL bilaterally (Mummery et al., 2000), which data from other methods in healthy participants and patient groups has shown to be a key region for the formation of coherent concepts (Lambon Ralph et al., 2010). This finding agrees with Sajjadi et al. (2017) who found that the atrophy patterns for SD were more easily distinguishable (high sensitivity and specificity) than the other forms of PPA.

Describing the symptomatology of PSA and PPA in terms of gradedly-different regions within multidimensional space has a number of potential clinical implications. First, this approach allows us to begin to determine both the range and type of variations that are associated with each of the pre-existing clinical labels, rather than reserving the use of each diagnostic label to a single, invariant prototypical pattern. Secondly, by extension, it also allows us to establish when and why certain subtypes of PPA are most likely to be confused with each other, and the same for subtypes of PSA. Thirdly, it provides a single unifying framework within which both established and “mixed” aphasias can be considered.

Fourth, future clinical research can explore whether considering the phenotype variations along continuous, dimensions (i.e., a transdiagnostic approach) rather than categorical systems might reveal clearer relationships between phenotype and atrophy, pathology or genetic markers. For example, past work in PSA has shown that utilising raw individual test scores or PSA categories leads to undifferentiated lesion correlates reflecting the whole MCA territory rather than specific subregions. When the same lesion-mapping analyses are repeated using the PCA-derived dimensions then much more discrete and interpretable subregions are revealed (cf. Butler et al., 2014). These clearer symptom-lesion maps can then be inverted in order to generate lesion-based diagnostics and prediction models (Halai et al., 2018a). Finally, taking this multidimensional approach could inform a transdiagnostic selection process for treatment, therapy or clinical trials; in order to select a group of patients with relatively homogeneous behavioural symptoms, one could select patients who occupy a shared region of the multidimensional space (thereby sharing symptomatology across the core language systems captured by the dimensions), irrespective of their clinical diagnosis. Indeed, the importance of a transdiagnostic approach has been highlighted for Frontotemporal Lobar Degeneration (FTLD) with regards to shared apathy and impulsivity symptomatology (Lansdall et al., 2017, Passamonti et al., 2018, Lansdall et al., 2019).

In conclusion, we have shown that the internal structure of variation in PSA and PPA, in isolation or in a unified framework, can be captured with the same four underlying language-cognitive dimensions. Furthermore, semantic dementia appears to represent a robust
diagnostic category, whilst patients with other forms of PPA and PSA might be better described in terms of their gradedly-different positions along these four principal language-cognitive dimensions.
Supplementary Figure 2.4 – Principal components extracted for post-stroke aphasia.

Connected speech was elicited using the ‘cookie theft’ picture description task from the BDAE (Kaplan, 1983). Connected speech was quantified in the parameters: total number of words/tokens, type-token ratio, mean length per utterance, and words-per-minute. Comprehension of grammar was assessed with the spoken sentence comprehension task from the Comprehensive Aphasia Test (Swinburn et al., 2004). Subtests from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) battery (Kay et
al., 1992) were used to assess auditory discrimination (words and non-words minimal pairs), and repetition of words and non-words (immediately and after a delay). Semantic abilities were tested with the Cambridge Semantic Battery (CSB) (Bozeat et al., 2000), including the spoken and written version of the word-to-picture matching task, the 64 item naming test, and the Camel and Cactus Test (picture version). Additionally, the Boston Naming Test (BNT) (Goodglass et al., 1983) and the written 96-trial Synonym Judgement test (Jefferies et al., 2009) were included as these are more sensitive to mild semantic deficits. Non-language measures of attention and executive function included digit span (forward and backwards) (Wechsler, 1981), the Brixton Spatial Rule Anticipation Task (Burgess and Shallice, 1997), and Raven's Coloured Progressive Matrices (Raven and Court, 1962).
Supplementary Figure 2.5 – Principal components extracted for primary progressive aphasia.

Connected speech samples were elicited via picture description task from the Comprehensive Aphasia Test (Swinburn et al., 2004) and quantified in the parameters: total number of words, mean length of utterance (MLU), and words-per-minute (WPM).
Comprehension of grammar was assessed through the Test for Reception of Grammar (TROG) (Bishop, 1989) and Sentence Comprehension Test (auditory and visual presentations – SECTA/SECTV) (Billette et al., 2015). The Make A Sentence Test (MAST) (Billette et al., 2015) was used to measure grammatical ability in sentence production. Ability to repeat was tested for words, non-words and sentences (Sajjadi et al., 2012a). Repetition of words was also measured in the Repeat and Point test (Hodges et al., 2008). Semantic ability was measured through the 64-item naming test and the picture version of the Camel and Cactus Test of semantic association knowledge (CCT), both from the Cambridge Semantic Battery (CSB) (Bozeat et al., 2000). The Point aspect of the Repeat and Point test (Hodges et al., 2008) also measured object knowledge. Phonological skill was assessed through non-word minimal pairs from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) (Kay et al., 1992) for phonological perception, and forward letter span (Sajjadi et al., 2012a) for measuring the capacity of the phonological loop. Addenbrooke’s Cognitive Examination – Revised (ACE-R) (Moshi et al., 2006) was included as a general assessment which tapped multiple domains of cognition and language. Digit span (Wechsler, 1981) (forwards and backwards) was used to measure attention and executive function, alongside subsections A and B of the Delis-Kaplan executive function system (D-KEFS) Trail Making Test (Delis et al., 2001). Visuospatial skills were tested using the cube analysis subsection of the Visual Object and Space Perception (VOSP) battery (Warrington and James, 1991). The Rey-Osterrieth Complex Figure (Rey, 1941) was also used to measure visuospatial ability in the direct copying of the figure, and non-verbal memory in the later recall of the figure at a delay. Sajjadi et al. (2012a) also developed tests for oro-buccal (mouth/cheek) and limb praxis, including transitive (meaningful) and intransitive (meaningless) gestures.
Supplementary Figure 2.6 – Principal components extracted for the unified principal component analysis of primary progressive aphasia and post-stroke aphasia.
Supplementary Table 2.3 – Iterative sweep through Unified PCA multidimensional space to find cut-offs giving optimal diagnostic selectivity.

<table>
<thead>
<tr>
<th>Group</th>
<th>Subtype</th>
<th>Phonology</th>
<th>Semantics</th>
<th>Visuo-executive</th>
<th>Speech production</th>
<th>d prime</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA</td>
<td>Anomia</td>
<td>&gt; -1.14</td>
<td>&gt; 0.38</td>
<td>&gt; -1.49</td>
<td>&lt; 1.10</td>
<td>2.64</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Broca</td>
<td>&lt; -0.02</td>
<td>&gt; 0.33</td>
<td>&gt; 0.92</td>
<td>&lt; -0.98</td>
<td>2.37</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>Conduction</td>
<td>&lt; -1.22</td>
<td>&gt; -0.25</td>
<td>&gt; 0.47</td>
<td>&gt; -0.88</td>
<td>3.11</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>&lt; -1.62</td>
<td>&lt; -0.88</td>
<td>&gt; 0.38</td>
<td>&lt; 0.22</td>
<td>3.40</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Mixed non-fluent</td>
<td>&lt; 1.23</td>
<td>&lt; 0.39</td>
<td>&gt; 0.53</td>
<td>&lt; -1.76</td>
<td>2.37</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>TMA</td>
<td>&gt; 0.68</td>
<td>&gt; 0.54</td>
<td>&gt; 0.53</td>
<td>&lt; -1.70</td>
<td>2.17</td>
<td>0.0288</td>
</tr>
<tr>
<td>PPA</td>
<td>lvPPA</td>
<td>&gt; -0.19</td>
<td>&gt; 0.28</td>
<td>&lt; -0.59</td>
<td>&gt; 0.49</td>
<td>3.38</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>mPPA</td>
<td>&lt; 1.51</td>
<td>&lt; 0.73</td>
<td>&lt; 0.54</td>
<td>&lt; 1.35</td>
<td>2.10</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>PNFA</td>
<td>&lt; 1.37</td>
<td>&gt; 0.07</td>
<td>&lt; 0.73</td>
<td>&lt; 1.83</td>
<td>2.03</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>&gt; 0.29</td>
<td>&lt; 0.98</td>
<td>&gt; -0.14</td>
<td>&gt; -0.52</td>
<td>4.46</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Cut-off values are presented as factor scores on each extracted dimension. Abbreviations: PSA – post-stroke aphasia; PPA – primary progressive aphasia; TMA – transcortical motor aphasia; lvPPA – logopenic variant PPA; mPPA – mixed PPA; PNFA – progressive non-fluent aphasia; SD – semantic dementia.

Diagnostic cut-off values were derived from the iterative sweep through the Unified principal component analysis multidimensional space. The cut-off values (columns 3-6) are factor scores on the four dimensions. This allowed $d'$ values (column 7) associated with each combination of cut-off values to be extracted. To establish the likelihood of achieving these $d'$ values by chance, diagnostic group-membership was randomised within aetiology and $d'$ re-calculated over 10,000 iterations to give a distribution of $d'$ values, and to calculate the $p$ value (column 8) associated with each of the optimum diagnostic cut-off combinations.
Supplementary Table 2.4 – Distribution of misclassifications between clinical and data-driven diagnostic PSA groups.

<table>
<thead>
<tr>
<th>Clinical diagnostic groups (N)</th>
<th>Hits</th>
<th>Misclassifications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Anomia</td>
</tr>
<tr>
<td>Anomia (15)</td>
<td>100.0</td>
<td>/</td>
</tr>
<tr>
<td>Broca (5)</td>
<td>60.0</td>
<td>60.0</td>
</tr>
<tr>
<td>Conduction (3)</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Global (5)</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Mixed non-fluent (5)</td>
<td>60.0</td>
<td>0.0</td>
</tr>
<tr>
<td>TMA (1)</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Data expressed as percentages of the total number of cases in each clinical diagnostic group. Abbreviations: PSA – post-stroke aphasia; TMA – transcortical motor aphasia.

Amongst the PSA cases, data-driven diagnostic criteria for three subtypes achieved a perfect hit rate and no misclassifications: global aphasia, conduction aphasia, and TMA subtypes. The data-driven diagnostic criteria for global aphasia and conduction aphasia had the highest d’ values (global: $d' = 3.40, p < .001$; conduction: $d' = 3.11, p < .001$), whilst the TMA criteria had a lower d’ value ($d' = 2.17, p = 0.029$). These d’ values are lower than that for SD due to the smaller sample sizes in the PSA cohort, which affect the adapted d’ calculation (Macmillan and Kaplan, 1985). The anomia data-driven diagnostic group had a perfect hit rate but misclassifications of Broca’s aphasia and TMA cases resulting in a lower d’ value ($d' = 2.64, p < .001$). The data-driven diagnostic criteria for Broca’s aphasia incurred no misclassifications of other subtypes but failed to correctly classify all the Broca’s cases (Hit rates around 60%), resulting in lower d’ values (Broca: $d' = 2.37, p = .001$). This shows that, like PNFA, even the optimal data-driven diagnostic criteria for Broca’s aphasia and were insufficiently selective, suggesting that this subtype of PSA does not meet the assumptions of a true category. The mixed non-fluent aphasia group represents cases who did not meet the criteria for a single proposed subtype of PSA. The data-driven diagnostic criteria for this group ($d' = 2.37, p = .002$) were poorly selective, as they failed to capture all the mixed non-fluent PSA cases. This reinforces the label applied to these cases.
Chapter 3 - Assessing the validity of the novel Mini-Linguistic State Examination in post-stroke aphasia

Ruth U. Ingram¹, Blanca De Dios Perez⁵, Ajay D. Halai², Nikil Patel⁴, Katie Peterson³,
Gorana Pobric¹,
Karalyn Patterson²,³, Peter Garrard⁴ & Matthew A. Lambon Ralph²

¹Division of Neuroscience and Experimental Psychology,
School of Biological Sciences, University of Manchester, UK
²MRC Cognition & Brain Sciences Unit, University of Cambridge, UK
³Department of Clinical Neurosciences, University of Cambridge, UK
⁴Centre for Clinical Neuroscience, St George's University of London, UK
⁵University of Nottingham, UK

Statement of contribution
Ruth Ingram analysed all the data presented in this chapter. Blanca De Dios Perez collected
the behavioural data on the chronic post-stroke aphasia cohort. Nikil Patel and Katie
Peterson collected the behavioural data on the healthy controls. Matt Lambon Ralph, Ajay
Halai, and Gorana Pobric provided guidance and support throughout the analysis, and
feedback on drafts of the write up.
Abstract

Aphasia refers to impairment of the ability to comprehend and formulate language which manifests as difficulties across multiple modalities of language use. Aphasia can be caused by stroke (post-stroke aphasia (PSA) or neurodegeneration (PPA). Despite the similarity of symptoms and nomenclature in subtypes of PSA and PPA, the lack of a standardised linguistic tool which has been validated for aphasia across these aetiologies has likely contributed to a lack of systematic comparisons. The Mini-Linguistic State Examination (MLSE) is a novel clinical assessment tool which is in development to address this gap. This study conducted the first formal validation of the MLSE and compared it to the Boston Diagnostic Aphasia Examination (BDAE) (Kaplan, 1983), and also to a range of comprehensive, in-depth language assessments targeting the same linguistic domains as the MLSE subtests. We conducted this comparison and validation using a cohort of PSA for two key reasons: (1) these established aphasia tests have been designed and validated for PSA, and (2) it revealed the profile of MLSE performance across mild and moderate-to-severe PSA subtypes, laying the groundwork for future comparisons of PSA and PPA using this novel tool. The validity of the MLSE was assessed in terms of: (1) convergence of the subtype classifications derived from MLSE test scores vs. BDAE test scores; (2) sensitivity of the MLSE compared to the in-depth tests in terms of (a) sensitivity to mild deficits, and (b) sensitivity to levels of impairment; and (3) convergent validity of the MLSE subtests compared with the in-depth tests. Overall, we found that the MLSE is hypersensitive to the presence of a language impairment in PSA and that many of the MLSE subtests show good convergent validity. Furthermore, MLSE subtest performance was able to differentiate milder cases from cases with moderate to severe impairments. Future research can capitalise on this transdiagnostic, sensitive language assessment tool to compare brain-behavioural relationships in aphasia across aetiologies.
Introduction

Aphasia is an impairment of the ability to comprehend and formulate language following acquired brain damage, which manifests as difficulties across multiple modalities of language use (e.g., reading, auditory comprehension, expressive language) (Rosenbek et al., 1989). Aphasia can arise due to multiple aetiologies of brain injury. Left-hemisphere stroke is a common cause of aphasia (Berthier, 2005), which we will refer to as post-stroke aphasia (PSA). Primary progressive aphasia (PPA) is the term given to aphasia which is the most salient symptom of neurodegenerative disease (Mesulam, 2001a). Despite the similarity of symptoms and nomenclature in subtypes of PSA and PPA, there have been few detailed direct comparisons across the full ranges of PSA and PPA. The lack of a standardised linguistic tool which has been validated for aphasia across these aetiologies has likely contributed to this paucity of systematic comparisons. The Mini-Linguistic State Examination (MLSE) is a novel clinical assessment tool which is in development to address this gap. This study conducted the first formal validation of the MLSE and compared it to established aphasia tests. We compared the performance of people with PSA on the established tests and the MLSE for two key reasons: (1) the established aphasia tests have been designed and validated for PSA, and (2) it enabled us to establish the profiles of MLSE performance in mild and moderate-to-severe PSA subtypes, laying the groundwork for future comparisons of PSA and PPA using this novel tool.

The MLSE is designed to assess a range of linguistic impairments which are known to be impacted in PPA (Gorno-Tempini et al., 2011) but also in PSA (Berthier, 2005, Berthier et al., 2014) and are compromised in other neurodegenerative disorders such as corticobasal syndrome and progressive supranuclear palsy (Peterson et al., 2019). Briefly, the MLSE assesses naming, repetition (syllables, words, non-words and sentences), sentence comprehension, semantic memory (word-picture matching and semantic association), reading (words and non-words), writing, and picture description. These subtests were chosen to capture the linguistic domains which are used for differential diagnosis of PPA subtypes (Gorno-Tempini et al., 2011). Impaired word retrieval and sentence repetition characterise the logopenic subtype (Gorno-Tempini et al., 2008). Agrammatism in language production, and effortful speech characterise the non-fluent/agrammatic variant (Nestor et al., 2003, Gorno-Tempini et al., 2004). Anomia and impaired single-word comprehension are characteristic of the semantic variant, also known as semantic dementia (Snowden et al., 1989, Lambon Ralph and Patterson, 2008).

These PPA subtypes are broadly reflected in subtypes of PSA (Grossman and Irwin, 2018). For example, progressive non-fluent aphasia appears to resemble Broca’s aphasia, as both display effortful, dysfluent speech and impaired grammatical competence (Thompson et al., 1997) (though see Patterson et al., 2006a). Logopenic progressive aphasia resembles conduction aphasia (Rohrer et al., 2013, Cummings, 2008), though they have not been systematically compared. Filley et al. (2006) reported a case of progressive anomia with left
temporoparietal hypometabolism (consistent with logopenic progressive aphasia) who met the Boston Diagnostic Aphasia Examination diagnostic criteria for conduction aphasia (Kaplan, 1983), i.e., a deficit in ability to repeat despite intact auditory comprehension and relatively fluent speech (Ardila, 2010b, Benson et al., 1973). Semantic dementia might be comparable to Wernicke’s aphasia (Ogar et al., 2011), which is also characterised by impaired comprehension in the context of fluent speech. Wernicke’s aphasia presents as impaired auditory comprehension such that connected speech, although fluent, contains meaningless words, sometimes with irregular syntactic structure (Goodglass et al., 1983, Kertesz et al., 1979). Table 3.1 shows the diagnostic classification system for other subtypes of PSA, in terms of varying impairments which can also be captured by ability to repeat, comprehend, and speak fluently.

**Table 3.1 – Subtypes of post-stroke aphasia according to the Boston Diagnostic Aphasia Examination classification system.**

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Fluency</th>
<th>Comprehension</th>
<th>Repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anomic aphasia</strong></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>Conduction aphasia</strong></td>
<td>✔️</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td><strong>Transcortical sensory aphasia</strong></td>
<td>✔️</td>
<td>✗</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>Wernicke’s aphasia</strong></td>
<td>✔️</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td><strong>Transcortical motor aphasia</strong></td>
<td>✗</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>Broca’s aphasia</strong></td>
<td>✗</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>Mixed transcortical atrophy</strong></td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Global aphasia</strong></td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

Ticks indicate relatively spared domains whilst crosses are impaired.

Thus, exploring the performance of people with PSA on the MLSE serves two purposes: (1) it shows whether the MLSE subtests are sensitive to the types of deficit which will be important across aetiologies of aphasia, and (2) it is required in order to validate the MLSE against established tests which have been validated for PSA. By ‘established’ tests, we refer to tests which have been extensively validated and are used and trusted by researchers and clinicians alike for assessing aphasia. In line with the long history of assessing language impairment in PSA (Tesak and Code, 2008), there are many established language assessment tools for PSA. These include the Boston Diagnostic Aphasia Examination (BDAE) (Kaplan, 1983), Western Aphasia Battery (Kertesz, 1982), and Comprehensive Aphasia Test (Swinburn et al., 2004). We compared the MLSE to the BDAE.
The validity of novel tests needs to be established in terms of sensitivity to the impairments assessed and convergence of the constructs measured (Ivanova and Hallowell, 2013). Firstly, sensitivity to mild deficits is particularly important in the context of differential diagnosis of mild cases of aphasia from cognitive changes in normal aging (Ross and Wertz, 2004). This is vital for a number of reasons, such as ensuring that individuals with mild aphasia are not mistakenly discharged or withdrawn from clinical services due to lack of apparent impairment (Johnson et al., 1998). This is important because often even mild aphasia is worsened by tiredness/busy environments outside of the clinical setting and so still significantly impacts activities of daily living (Gialanella et al., 2016). Additionally, an accurate representation of the epidemiology of aphasia due to stroke requires identification of the full spectrum of language impairments (Flowers et al., 2016). Secondly, sensitivity to levels of severity in language impairment is important for indicating which clinical pathways are appropriate/necessary (Shrubsole et al., 2017) (e.g., speech and language therapy, or more intensive support), since the level of impairment might inform expectations of response to therapy (Watila and Balarabe, 2015)). Novel tests also need to display construct validity (Ivanova and Hallowell, 2013). This form of validity indicates whether the same impairments are captured by tests proposing to measure the same language abilities. Construct validity is comprised of convergent validity and discriminant validity (Mitrushina et al., 2005), which are both assessed through correlations. Convergent validity is evident if the novel test strongly correlates with an established test which proposes to measure the same language ability. Discriminant validity is implied when tests proposing to measure different language abilities are not strongly correlated.

One aspect of sensitivity and convergent validity which can be assessed together is whether the information captured by the MLSE results in the same classifications of PSA subtypes (i.e., whether a person receives the same subtype diagnosis) as the information captured by an established diagnostic test. The BDAE is an appropriate comparison for this form of validity because, like the MLSE, it is designed to facilitate diagnosis of aphasic subtypes, intended for quick and efficient use in the clinic, and comprised of subtests to assess a range of linguistic abilities with only a few items. The BDAE delineates eight subtypes of PSA based on impairments in three domains: repetition, comprehension, and fluency. Using this classification system we compared the subtype classifications derived for the same people with PSA using the MLSE and the BDAE. This analysis revealed the convergence and divergence of subtype classifications derived from the MLSE and BDAE at the individual level, thus allowing us to explore which subtypes of PSA were most often given different classifications by the MLSE and BDAE.

However, the BDAE is not necessarily the most appropriate base of comparison for assessing the other forms of validity, for two reasons. Firstly, the BDAE does not assess all the linguistic domains measured by the MLSE (the MLSE is being developed to address the lack of established tools which assess all these domains simultaneously). Secondly, the
brevity of the BDAE subtests comes at the expense of exhaustive, detailed assessment of the linguistic domains; established tests with more items, and/or items of graded difficulty, have higher acuity for capturing subtle and nuanced language impairments. Therefore, we assessed the sensitivity and convergent validity of the MLSE compared to a comprehensive range of linguistic tests with greater depth than the BDAE subtests. These in-depth tests were chosen to match the MLSE subtests as closely as possible, and so provided a more fine-grained comparison for the linguistic abilities captured by the MLSE. Thus, we contrasted each MLSE subtest with a matched in-depth test, in order to (a) compare sensitivity to mild deficits, (b) compare sensitivity to levels of impairment, and (c) establish the convergent validity of the MLSE subtest.

Method

Participants

Twenty-one people with chronic post-stroke aphasia were prospectively recruited from community groups and speech and language therapy services in the North West of England. Patients were included if they reported a single left hemisphere stroke at least 12 months prior to assessment and were native English speakers. Patients were recruited at all levels of severity, thus this cohort included a heterogeneous collection of aphasic phenotypes, classified using the Boston Diagnostic Aphasia Examination (Kaplan, 1983) as: anomic aphasia (N = 10), Broca’s aphasia (N = 3), conduction aphasia (N = 3), global aphasia (N = 2), mixed non-fluent aphasia (N = 2) and transcortical motor aphasia (N = 1).

Twenty-seven people were recruited as age- and education-matched controls from St George’s University, London, UK and the University of Cambridge, Cambridge, UK. Healthy controls were excluded from participation if they reported a history of neurological, psychological, or speech, language, and/or learning deficits. All healthy controls were native English speakers with normal or corrected to normal hearing and vision.

All participants gave informed consent under approval from respective local ethics committees.

Test batteries

MLSE

The MLSE comprises 13 subtests intended to capture variation in fundamental aspects of language production and comprehension. The MLSE subtests are: picture naming; repetition of syllables, words, non-words and sentences; sentence comprehension (I and II); semantic memory (spoken word-picture matching and semantic association); reading words with irregular sound-spelling associations (e.g., ‘mauve’) and reading non-words; procedural writing (‘explain to someone how to brush their teeth’); and spoken picture description (beach scene). Sentence comprehension was assessed through two subtests; sentence comprehension I involves participants listening to a sentence with complex syntactic structure, followed by a question relating to the sentence (e.g., ‘Joe was treated by Mary. ...
Who was the doctor?’) that is read by the experimenter. The participant answers verbally. Sentence comprehension II also involves a syntactically-complex sentence spoken by the experimenter, which the participant must match to a picture illustrating that sentence. The target picture is accompanied by 3 distractors which contain the same action/actors in different confirmations. Word repetition and spoken word-picture matching mimic each half of the ‘repeat and point’ test (Hodges et al., 2008), wherein the participant repeats a single word after the experimenter, then points to the picture corresponding to that word (from a selection including 5 distractors). The sentence repetition test was of graded difficulty, increasing from 6 to 15 words in length, and also contained sentences with an ‘unexpected’ conclusion (e.g., ‘the astronomer gazed at the flowers’).

The Mini-Linguistic State Examination is designed to be a brief yet comprehensive test for use in clinical settings, and as such short completion time is prioritised (the MLSE can be administered in 20 minutes on average, with 5 further minutes for calculating the scoring). Therefore, a small selection of items (difficult items or of graded difficulty) are included for each subtest (minimum of three, maximum of six).

**Matched battery**

We created a comprehensive test battery to compare against the MLSE, which we termed the Matched battery. The following in-depth linguistic tests were chosen to match each of the MLSE subtests:

- Naming was matched the Cambridge Semantic Battery 64-item naming test (Bozeat et al., 2000) (control data sourced from Garrard et al. (2001)) and the Boston Naming Test (Goodglass et al., 1983) (control data sourced from the manual).
- Spoken word-to-picture matching and semantic association were matched with the corresponding Cambridge Semantic Battery tests (Bozeat et al., 2000) (control data sourced from Garrard et al. (2001)).
- Word and non-word repetition were matched against the Psycholinguistic Assessments of Language Processing Abilities (Kay et al., 1992) immediate repetition tests 9 and 8, respectively (control data sourced from the manual, and Davies and Bose (2019), respectively).
- Sentence repetition was matched against a sentence repetition test designed for PPA with accompanying control data (Sajjadi, 2013)).
- Sentence comprehension involving a verbal response or selecting from picture stimuli were matched against the Sentence Comprehension Test - Auditory (SECTA (Billette et al., 2015)) and the Test for the Reception of Grammar (Bishop, 1989), respectively; control data for these matched tests was sourced from Billette et al. (2015) and Sajjadi (2013), respectively.
- Irregular word reading was matched with the 84 words with exceptional sound-spelling associations from the Surface List (Patterson and Hodges, 1992) (control data from N. Graham (personal communication, 23rd August 2019)).
• Non-word reading was matched with data from the PALPA8 non-word list.
• Writing was matched with the Comprehensive Aphasia Test written picture description assessment (Swinburn et al., 2004) (control data sourced from the manual).
• Spoken picture description was matched with a spoken picture description dataset (personal communication from R. Alyahya (12/08/2019), also published as a doctoral thesis (2019)).
• The syllabic repetition task did not have an in-depth matched equivalent.

Data analysis
Accuracy data were converted to percentages for all analyses. Statistical analysis was conducted in SPSS 25 (IBM statistics). Demographic details were compared for patients (as a whole group due to small sample sizes at the subtype-level) compared to healthy controls, using independent t-tests (age, education) and Chi-squared test (sex). The validity of the MLSE was assessed in three ways which are outlined below: (1) convergence of the subtype classifications derived from MLSE test scores vs. BDAE test scores; (2) sensitivity of the MLSE compared to the Matched test battery in terms of (a) sensitivity to mild deficits, and (b) sensitivity to levels of impairment; and (3) convergent validity of the MLSE subtests compared with the corresponding Matched tests.

Convergence of subtype classifications
All patients were first classified into diagnostic subtypes using performance on Boston Diagnostic Aphasia Examination (Kaplan, 1983) (Table 3.1). Fluency was assessed through picture description, particularly speech rate and number of words produced. Comprehension was assessed through word-to-picture matching. Repetition was assessed through word repetition and sentence repetition. Performance in the same three domains were then used to classify cases again using their MLSE data. Fluency was assessed through the number of elements produced in the picture description, plus the presence of articulatory or grammatical errors. Comprehension was assessed through spoken word-to-picture matching. Repetition was assessed through repetition of words, non-words, and sentences. Cut-offs 2 standard deviations below the control mean were used to signify impairment. We then quantified the percentage of cases for whom the BDAE-derived and MLSE-derived subtype classification agreed, for each subtype.

Sensitivity to mild deficits and levels of impairment
To explore the sensitivity of the MLSE subtests to mild impairments, we aimed to see how often a patient displaying an impairment (i.e., falling below control performance) on a detailed Matched test also displayed an impairment on the corresponding MLSE subtest (and vice versa). To classify impairment, cut-off scores were generated from control data (MLSE) or from published sources (Matched tests). A score of two standard deviations below the control mean was used as the cut-off for impairment. When an individual case fell
below the cut-off for the MLSE subtest but not the corresponding test from Matched battery, this was considered to show increased sensitivity to mild deficits in the MLSE subtest, and vice versa if they fell below the cut-off on the Matched test. In other words, this was taken to suggest that one test had detected the presence of an impairment whilst the other test had missed this impairment, and therefore that the first test appeared to be more sensitive. We quantified the percentage of cases whose impairments were ‘missed’ by either the MLSE test or corresponding Matched test. We compared the distribution of ‘missed’ cases by MLSE and Matched tests using Wilcoxon signed-rank test.

To investigate the sensitivity of the MLSE to levels of impairment, we assessed whether there were significant differences in the performance of cases defined as mildly impaired or moderately-to-severely impaired. We used Principal Component Analysis (PCA) to derive a single continuous factor which captured global severity by taking into account performance across the BDAE. PCA is a form of factor analysis which extracts a set of linearly uncorrelated variables (called components or factors) from a set of inter-correlated observations (Jain and Shandliya, 2013). By entering BDAE performance into PCA and extracting a single component (explaining 62.0% of variance in BDAE performance), we were able to generate a single variable capturing global aphasia severity. Using a median split of cases’ scores on this BDAE PCA-derived severity factor, we split the cases into Mild or Moderate-Severe groups. We also used a Matched test battery PCA-derived severity factor (explaining 60.5% of variance in Matched test performance) to split cases to see if this would be more informative. The Mild/Moderate-Severe median splits for the Matched severity factor and the BDAE severity factor were identical, so the BDAE PCA-derived severity factor split is reported here.

We conducted a 3 x 13 mixed Analysis of Variance (ANOVA) (3 groups, 13 MLSE subtests) to see if there were significant differences in performance across the MLSE subtests between healthy controls, Mild PSA and Moderate-Severe PSA. When the assumption of homogeneity of variance was violated (e.g., significant result of Mauchly’s test for sphericity or Levene’s test for homogeneity of variance) appropriate corrections were applied. Following a significant ANOVA result, we conducted pairwise multiple comparisons to explore which subtests and groups were significantly different. Specifically, we conducted 13 independent t-tests to compare MLSE subtests performance across each combination of the Group independent variable. Bonferroni correction for multiple comparisons was applied, with significant differences being considered at p < .05 after correction. Significant differences between patients in the Mild group and healthy controls were taken to indicate sensitivity of the MLSE subtests to mild impairment. Significant differences between the performance of the Mild group and the Moderate-severe group were taken to indicate sensitivity to levels of impairment.
Convergent validity of MLSE subtests

To investigate the convergent validity of the MLSE subtests, we calculated the Pearson correlation between each subtest and its corresponding test from the Matched battery. Based on the MLSE’s sensitivity to levels of impairment, we also assessed whether the correlations between MLSE subtests and Matched tests could be driven by general severity. For example, if more severe cases do worse because of a global impairment, and mild cases do better because of the opposite, and this difference in severity is maintained across all tests, then it would lead to all tests being correlated. We investigated this by correlating the MLSE subtests with the BDAE PCA-derived severity factor.

Specifically, we computed 91 Pearson correlation coefficients for every pairwise combination of MLSE subtests and also 13 correlations for each MLSE subtest with the BDAE PCA-derived severity factor. Bonferroni correction for multiple comparisons was applied, with significant differences being considered at $p < .05$ after correction. A strong significant positive correlation between MLSE and Matched tests was taken to indicate convergent validity, i.e., the two tests measure the same phenomenon. Strong significant positive correlations between an MLSE subtest and a non-target Matched test (i.e., one proposed to measure a very different language ability) were considered evidence of poor discriminant validity. Significant correlations of MLSE subtests with the severity factor indicate that the correlation is being driven by generalised severity.
Results

Demographics

The demographic details of the MLSE cohort are shown in Table 3.2. There was no significant difference in Age between healthy controls and patients (t(46) = -1.155, p = .254) but controls were had significantly more years of education (t(46) = 4.382, p < .001). There were no significant differences in sex distribution in healthy controls and patients (χ²(1) = .012, p = .912).

Table 3.2 – Demographic details of post-stroke aphasics and healthy controls.

<table>
<thead>
<tr>
<th>BDAE subtype</th>
<th>N (F)</th>
<th>Age</th>
<th>Education</th>
<th>Time post stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomia</td>
<td>10 (6)</td>
<td>60.5 (13.9)</td>
<td>11.8 (2.7)</td>
<td>6.1 (6.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[44 – 86]</td>
<td>[9 – 19]</td>
<td>[1.4 – 23.2]</td>
</tr>
<tr>
<td>Broca</td>
<td>3 (1)</td>
<td>63.7 (10.7)</td>
<td>13.0 (3.5)</td>
<td>6.1 (4.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[52 – 73]</td>
<td>[11 – 17]</td>
<td>[2.8 – 9.5]</td>
</tr>
<tr>
<td>Conduction</td>
<td>3 (1)</td>
<td>60.0 (12.1)</td>
<td>14.7 (3.2)</td>
<td>1.3 (0.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[46 – 67]</td>
<td>[11 – 17]</td>
<td>[1.1 – 1.8]</td>
</tr>
<tr>
<td>Global</td>
<td>2 (0)</td>
<td>60.0 (11.3)</td>
<td>11.5 (0.7)</td>
<td>5.1 (1.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[52 – 68]</td>
<td>[11 – 12]</td>
<td>[4.2 – 6.1]</td>
</tr>
<tr>
<td>Transcortical Motor Aphasia</td>
<td>1 (1)</td>
<td>73</td>
<td>11</td>
<td>3.8</td>
</tr>
<tr>
<td>Mixed non-fluent aphasia</td>
<td>2 (0)</td>
<td>69.0 (1.4)</td>
<td>11</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[68 – 70]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy Controls</td>
<td>27 (12)</td>
<td>58.6 (10.3)</td>
<td>15.9 (2.8)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[40 – 73]</td>
<td>[11 - 21]</td>
<td></td>
</tr>
</tbody>
</table>

Age, education, and time post-stroke are expressed in years (mean (standard deviation) [range]).

Convergence of subtype classifications

Table 3.3 shows the number of cases classified into each diagnostic subtype based on test scores on the BDAE and on the MLSE. Overall, even some mild cases performed more than 2 standard deviations below control norms (see Sensitivity to Levels of Impairment) on many of the MLSE subtests. For example, using test scores on the MLSE, 12/21 cases were classified as having global aphasia, compared to just 2/21 classified as this subtype by the BDAE test scores. These 12 cases included those which the BDAE classified as having anomic aphasia, Broca’s aphasia, Conduction aphasia, Mixed non-fluent aphasia, and transcortical motor aphasia (TMA). The 2 cases with anomic aphasia which the MLSE classified as global aphasia scored more than 2 standard deviations below control cut-offs on all tests, except one who scored 60% on picture description but made phonological errors in speech production. This shows that the MLSE subtests in question are very difficult, even
with only a relatively mild language impairment (so that an Anomic PSA case would score more than 2 standard deviations below control levels in terms of repetition, comprehension and fluency).

Sensitivity

Mild deficits

To establish whether the MSLE subtests are sensitive to mild impairments, we quantified the percentage of cases (of any subtype) which were impaired according to the MLSE control cut-off but considered spared according to the corresponding Matched test cut-off. This is illustrated in Figure 3.1. The bottom right quadrant of each scatterplot captures cases which meet this criterion, whilst the top left quadrant capture cases with the opposite scenario (i.e., ‘missed’ by the MLSE). Table 3.4 summarises the percentage of cases ‘missed’ by the MLSE and Matched test battery. Overall, there were more pairs where the Matched test missed a greater percentage of patients than the MLSE test, compared to the opposite pattern (though the proportion is non-significant (Wilcoxon rank test \( p = .118 \)). These results show that the MLSE is as sensitive to the presence of a mild language impairment as the more in-depth Matched tests.

Levels of impairment

Figure 3.2 shows the performance of patients and controls on the MLSE subtests. Patients were split into two groups labelled Mild and Moderate-severe, drawn from a median split on a PCA-derived Severity factor generated from performance on the BDAE. The ANOVA showed a significant difference across MLSE subtests (\( F(12,540) = 34.4, p < .001 \)), a significant difference across groups (\( F(2,45) = 128.2, p < .001 \)) and a significant interaction (\( F(24,540) = 11.5, p < .001 \)).

Patients from both levels of severity performed significantly worse than controls on many subtests of the MLSE. On the Semantic association test, both Mild and Moderate-severe patients performed as well as controls (\( p = 0.013 \)), showing that this test is insensitive to semantic impairments in this PSA cohort. Mild patients performed significantly worse than controls on all tests except Semantic association, Syllabic repetition and Sentence comprehension I and II. This further shows the sensitivity of many MLSE subtests to mild impairments. However, several tests did not show a significant difference between Mild and Moderate-Severe patients. These tests were: word repetition, semantic association, sentence comprehension II, non-word reading, sentence repetition and writing. Thus, these MLSE subtests appear to have poor sensitivity to levels of impairment in PSA.
Table 3.3 – Convergence of subtype classifications derived using the MLSE and BDAE.

<table>
<thead>
<tr>
<th>MLSE</th>
<th>ANOMIA</th>
<th>CONDUCTION</th>
<th>TSA</th>
<th>WERNICKE</th>
<th>TMA</th>
<th>BROCA</th>
<th>MIXED NON-FLUENT</th>
<th>GLOBAL</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDAE total</td>
<td>4</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>17%</td>
</tr>
<tr>
<td>ANOMIA</td>
<td></td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>30%</td>
</tr>
<tr>
<td>CONDUCTION</td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>TSA</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>WERNICKE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>TMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>BROCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>MIXED NON-FLUENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>GLOBAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>50%</td>
</tr>
<tr>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0%</td>
</tr>
</tbody>
</table>

Data are expressed as number of cases meeting criteria for each subtype, according to their scores on either the BDAE or the MLSE. Abbreviations: BDAE: Boston Diagnostic Aphasia Examination (Kaplan, 1983), MLSE: Mini-linguistic State Examination, TSA: transcortical sensory aphasia, TMA: transcortical motor aphasia.

Cases were classified into one of six possible stroke aphasia subtypes, based on performance on the BDAE (columns across the top). Then, patients were given a second classification using performance on the MLSE (rows on the left). Cells highlighted in grey show the distribution of classifications on each test battery. The Agreement cells in bold show the percentage of cases classified as the same subtype by both batteries. The ‘Agreement’ row shows the percentage of cases meeting the criteria for each subtype based on their BDAE data, who also met the criteria for the same subtype based on their MLSE data (e.g., 10 cases met the criteria for anomia based on their BDAE scores, and of these, 3 (30%) met the criteria for anomia based on their MLSE scores). The ‘Agreement’ column shows the percentage of cases meeting the criteria for each subtype based on their MLSE scores, who also met the criteria for the same subtype based on their BDAE data (e.g., 12 cases met the criteria for Global aphasia based on their MLSE scores, and of these, 2 (17%) also met the criteria for Global aphasia based on their BDAE scores).
Figure 3.1 – Sensitivity of the MLSE to mild deficits.

Performance of individual cases on a selection of MLSE subtests and corresponding Matched tests. PSA subtypes are shown in different colours. The orange and blue cut-off lines represent 2 standard deviations below control mean on the respective test. The Pearson correlation between MLSE subtest and Matched test is displayed in the upper right of each panel. Abbreviations: BNT – Boston Naming Test; sWPM – spoken word-picture matching, CCTp – Camel & Cactus Test, picture stimuli; PALPA - Psycholinguistic Assessments of Language Processing in Aphasia; TMA – transcortical motor aphasia.
Table 3.4 - Sensitivity of the MLSE to mild deficits.

<table>
<thead>
<tr>
<th>Test domain</th>
<th>Test/source</th>
<th>$r^2$</th>
<th>Sig.</th>
<th>% Spared</th>
<th>% Impaired</th>
<th>% missed Matched</th>
<th>% missed MLSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naming</strong></td>
<td>CSB 64-item naming</td>
<td>0.68</td>
<td>$p &lt; .001$</td>
<td>9.52</td>
<td>90.48</td>
<td>4.76</td>
<td>4.76</td>
</tr>
<tr>
<td></td>
<td>MLSE naming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Boston Naming test</td>
<td>0.70</td>
<td>$p &lt; .001$</td>
<td>19.05</td>
<td>80.95</td>
<td>9.52</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MLSE naming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Word picture matching</strong></td>
<td>CSB spoken WPM</td>
<td>0.33</td>
<td>$p = .007$</td>
<td>52.38</td>
<td>47.62</td>
<td>52.38</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MLSE spoken WPM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Semantic assoc.</strong></td>
<td>Camel &amp; Cactus test</td>
<td>0.58</td>
<td>$p &lt; .001$</td>
<td>80.95</td>
<td>19.05</td>
<td>80.95</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MLSE semantic assoc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sentence comprehension</strong></td>
<td>SECTA</td>
<td>0.53</td>
<td>$p &lt; .001$</td>
<td>14.29</td>
<td>85.71</td>
<td>0.00</td>
<td>14.29</td>
</tr>
<tr>
<td></td>
<td>MLSE sentence comp. I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test for Reception of Grammar</td>
<td>0.31</td>
<td>$p = .01$</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>19.05</td>
</tr>
<tr>
<td></td>
<td>MLSE sentence comp. II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Continued on the next page…]
<table>
<thead>
<tr>
<th>Test domain</th>
<th>Test/source</th>
<th>$r^2$</th>
<th>Sig.</th>
<th>% Spared</th>
<th>% Impaired</th>
<th>% missed Matched</th>
<th>% missed MLSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PALPA9 – word repetition</td>
<td>0.58</td>
<td>$p &lt; .001$</td>
<td>42.86</td>
<td>57.14</td>
<td>19.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MLSE word repetition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PALPA8 – non-word repetition</td>
<td>0.61</td>
<td>$p &lt; .001$</td>
<td>14.29</td>
<td>85.71</td>
<td>14.29</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MLSE non-word repetition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sajjadi et al. (2013)</td>
<td>0.53</td>
<td>$p &lt; .001$</td>
<td>4.76</td>
<td>95.24</td>
<td>4.76</td>
<td>9.52</td>
</tr>
<tr>
<td></td>
<td>MLSE sentences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surface List</td>
<td>0.37</td>
<td>$p = .004$</td>
<td>28.57</td>
<td>71.43</td>
<td>23.81</td>
<td>4.76</td>
</tr>
<tr>
<td></td>
<td>MLSE irregular word reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PALPA8 non-word reading *</td>
<td>0.89</td>
<td>$p &lt; .001$</td>
<td>14.29</td>
<td>85.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLSE non-word reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAT written picture description</td>
<td>0.79</td>
<td>$p &lt; .001$</td>
<td>9.52</td>
<td>90.48</td>
<td>4.76</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MLSE writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alyahya (2019) ‘cookie theft’</td>
<td>0.22</td>
<td>$p = .031$</td>
<td>33.33</td>
<td>66.67</td>
<td>19.05</td>
<td>14.29</td>
</tr>
<tr>
<td></td>
<td>MLSE ‘beach scene’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* We did not have control norms for non-word reading (PALPA8) therefore the % cases missed by either the PALPA8 or the MLSE non-word reading subtest could not be calculated. Abbreviations: CSB – Cambridge Semantic Battery; WPM – word-picture matching, SECTA – Sentence Comprehension Test Auditory presentation; PALPA - Psycholinguistic Assessments of Language Processing in Aphasia; CAT – Comprehensive Aphasia Test.

This table shows the strength and significance of correlations between the MSLE subtests and corresponding Matched tests. The “% Spared” and “% impaired” columns, show the percentage of cases which fell above or below the control cut-off on each test. The “% missed” columns show the percentage of cases whose impairments were ‘missed’ by either the MLSE test or corresponding Matched test.
Figure 3.2 – Sensitivity of the MLSE to levels of impairment.

This figure shows accuracy across MLSE subtests for patients and controls, which was analysed using 3 x 13 Mixed ANOVA. Patients were split into Mild or Moderate-Severe groups based on a data-driven global severity factor. Data presented as mean ± standard error. Significant differences between Mild and Moderate-severe patient groups are denoted by: * p < .05, ** p < .01, *** p < .001 (Bonferroni corrected for multiple comparisons). Significant differences between patients and controls are denoted by the grey shaded bars. Grey shaded bars show non-significant differences, so bars without shading indicate p < .05 (also corrected for multiple comparisons). Abbreviations: sWPM – spoken word-picture matching.
Convergent validity of MLSE subtests

Correlational analyses were used to assess the convergent validity of the MLSE subtests. Table 3.5 shows the correlation matrix between the MLSE subtests and the corresponding tests from the Matched test battery, and the BDAE PCA-derived severity factor. The MLSE Naming subtest showed strong correlations with the Cambridge Semantic Battery 64-item naming test and Boston Naming test, but also with repetition and irregular word reading tests, and the severity factor, suggesting these correlations are related to generalised aphasia severity. The spoken word-picture matching MLSE subtest was strongly and significantly correlated only with non-word repetition. The Semantic association subtest was only significantly correlated with two Matched semantic tests, spoken word-picture matching and the Camel & Cactus test of semantic association. Word repetition on the MLSE was strongly and significantly correlated with word repetition, but also repetition of non-words and sentences, Cambridge 64-item naming, and the severity factor. The latter may be driving these correlations through generalised severity. In contrast, non-word repetition only correlated with the corresponding Matched non-word repetition test. Sentence repetition on the MLSE was correlated with non-word repetition and the matched sentence repetition test. The MLSE sentence comprehension I (which involves auditory working memory demands and production of a verbal response) subtest was correlated with 64-item naming, and repetition and reading of non-words, and the severity factor. In contrast, the sentence comprehension II test (which involves selecting a target picture to match the heard sentence) did not significantly correlate with any Matched test, or the severity factor. Irregular word reading correlated with 64-item naming, and repetition of words and non-words but not the other reading tests. This is in contrast to non-word reading which correlated with the matched non-word reading test, and also the writing test and severity factor. Similarly, the MSLE writing subtest was correlated with the Matched tests for both writing and non-word reading. Finally, picture description on the MLSE was strongly and significantly correlated with the Boston naming test, non-word repetition and the severity factor. Overall, these results confirm the convergent validity of the MLSE subtests with a few exceptions which show poor discriminant validity possibly due to correlations being driven by generalised aphasia severity rather than selective linguistic impairments.
### Table 3.5 - Convergent validity of the MLSE subtests.

<table>
<thead>
<tr>
<th>MLSE subtests</th>
<th>Matched battery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CSB naming</td>
</tr>
<tr>
<td>Naming</td>
<td>0.82</td>
</tr>
<tr>
<td>Spoken word-picture matching</td>
<td>0.67</td>
</tr>
<tr>
<td>Semantic association</td>
<td>0.48</td>
</tr>
<tr>
<td>Word repetition</td>
<td>0.71</td>
</tr>
<tr>
<td>Non-word repetition</td>
<td>0.63</td>
</tr>
<tr>
<td>Sentence repetition</td>
<td>0.43</td>
</tr>
<tr>
<td>Sentence comprehension I</td>
<td>0.75</td>
</tr>
<tr>
<td>Sentence comprehension II</td>
<td>0.10</td>
</tr>
<tr>
<td>Reading – irregular words</td>
<td>0.84</td>
</tr>
<tr>
<td>Reading - non-words</td>
<td>0.58</td>
</tr>
<tr>
<td>Writing</td>
<td>0.54</td>
</tr>
<tr>
<td>Picture description</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Values represent the Pearson correlation coefficients between pairs of tests, with darker background shades indicating stronger correlation. Cells in bold indicate significant correlations (p < .05, corrected for multiple comparisons). Cells/groups of cells with a thick box border indicate tests which are matched for task demands (as closely as possible) or test the same domain (e.g., all repetition tests). The far-right column in red shows correlations with the BDAE PCA-derived Severity factor. Abbreviations: CSB – Cambridge Semantic Battery; BNT – Boston Naming Test; sWPM – spoken word-picture matching, CCTp – Camel & Cactus Test, picture stimuli; SECTA – Sentence Comprehension Test Auditory presentation; TROG – Test for Reception of Grammar.
Discussion

This study aimed to assess the convergent validity and sensitivity of the novel Mini-Linguistic State Examination (MLSE). This was achieved by comparing the MLSE against an established aphasia battery, the Boston Diagnostic Aphasia Examination, and a battery of in-depth corresponding tests. Overall, we found that the MLSE is highly sensitive to the presence of a language impairment in PSA.

Using the cut-offs for impairment derived from control data and the diagnostic process of the BDAE diagnostic classification system, the MLSE classified most of the cases as being impaired in fluency, repetition and comprehension, thereby being classified as having global aphasia. Several of these cases were classified as anomic based on performance on the BDAE. This contradictory classification suggests that the MLSE subtests are more sensitive to mild language impairments than the BDAE. Furthermore, the fact that the MLSE detected impairments beyond impaired naming (the defining symptom of anomic aphasia Goodglass and Wingfield, 1997)) highlights the heterogeneity within the category of anomic aphasia, which can be detected with sufficiently sensitive tests. Anomia can be caused by lesions in various cortical and subcortical left-hemisphere regions (Kertesz, 1979), meaning that the underlying cause of the retrieval deficit is highly variable. This means that patients with ‘anomia’ are a highly heterogeneous group (Maher and Raymer, 2004). Our finding suggests that the MLSE subtests were able to detect subtle language-cognitive impairments underlying, or additional to, the word retrieval deficit in these patients.

Previous comparisons of subtype classifications derived using different established aphasia tools have also demonstrated inconsistencies within individual cases (Crary et al., 1992). For example, Wertz et al. (1984) compared the subtype classifications for the same patients derived from test scores on the BDAE and the Western Aphasia Battery (WAB, Kertesz, 1982). Finding poor agreement in classification, with just 27% of individual cases receiving the same subtype diagnosis on these batteries, they concluded that this was due to many cases being ‘unclassifiable’ on the BDAE, whilst the WAB was generally able to classify these cases. In our study, we did not include the BDAE subtypes of ‘isolation’ aphasia, or the ‘unclassifiable’ bin. Our intent was to evaluate the agreement between the BDAE- and MLSE-derived classifications in the situation where a classification was forced, so they could be better compared.

The statistical comparison of MLSE performance across healthy controls, Mild patients, and Moderate-severe patients also showed sensitivity to mild deficits. Mild patients performed significantly worse than healthy controls on all subtests except semantic association, syllabic repetition, and sentence comprehension I and II. Semantic association appears to be particularly insensitive because Mild patients made no errors (100% accuracy) on this subtest, and even Moderate-severe cases were not significantly different from healthy controls on this test. The spoken word-picture matching and naming tests show that the Mild group do display semantic impairments, which the semantic association task appears not to
capture. Furthermore, when comparing how often the MLSE subtests and their corresponding Matched tests disagreed on whether a case was impaired relative to control cut-offs, the Matched tests were equally sensitive to mild impairments; the proportion of ‘missed’ cases (i.e., the case fell above control cut-offs for the Matched test but below the control cut-off for the MLSE subtest) was not statistically significantly different from the reverse. Part of this sensitivity to mild deficits is likely related to the fact that controls score consistently high or at ceiling for most of the MLSE subtests. This shows that the items chosen in the MLSE subtests are uniquely difficult for people with aphasia, not just difficult for someone of any ability. With control performance being at ceiling (i.e., no standard deviation) for 2 subtests (semantic association and spoken word-picture matching), this does inflate the number of cases who are classed as impaired because it only takes a single error (on very few items) to be considered impaired. These difficult items are likely driving the sensitivity of the MLSE to mild impairments.

In terms of sensitivity to levels of impairment, Mild cases performed significantly better on seven of the MLSE subtests compared to Moderate-severe cases. This effect could have been abolished by floor effects given the hypersensitivity to mild impairments due to difficult test items. Instead, it was possible to distinguish the Mild and Moderate-Severe groups in terms of MLSE subtest scores.

We found evidence of convergent validity for several MLSE subtests in comparison with corresponding in-depth tests from the Matched battery. Although insensitive to impairments relative to controls, performance of PSA cases on the semantic association MLSE subtest was significantly correlated with the Matched tests also measuring semantic knowledge. The MLSE tests measuring repetition of non-words and sentences also showed convergent validity with corresponding Matched tests. Non-word reading and writing showed convergent validity, but we also cross-correlated (i.e., MLSE non-word reading was correlated with the Matched writing test, and the MLSE writing test was correlated with the Matched non-word reading test), suggesting shared variance in these linguistic domains. None of the aforementioned MLSE subtests showed a significant correlation with the BDAE PCA-derived severity factor, suggesting that these correlations reflect convergent validity of the constructs being measured rather than underlying variance in generalised aphasia severity. In contrast, other MLSE subtests showed somewhat poor discriminant validity (i.e., significantly correlations with non-matched tests, such as naming with sentence repetition). In tests with poor discriminant validity which were not significantly correlated with the severity factor (semantic association, sentence repetition, non-word reading and writing), these off-target correlations were weaker than the on-target correlations. Wilson et al. (2018) also found off-target correlations to be generally less than their on-target correlations for the development of the Quick Aphasia Battery (QAB). In MLSE subtests which showed a significant correlation with the severity factor, these off-target correlations were occasionally stronger than the on-target correlations (e.g., the naming subtest correlated with non-word repetition...
at \( r = 0.9 \), and with the Matched naming tests at 0.82 and 0.8 respectively). This result shows that variation in generalised disease severity is contributing to relationships between several linguistic domains on the MLSE and Matched tests. Overall, in so far as a single subtest can be a pure measure of one aspect of language, the pairs of tests with selectively very high correlations (e.g., \( r^2 > .80 \)) suggest good convergent validity.

A limitation of the hypersensitivity of the MLSE to the presence of a language impairment is that it potentially reduces the ability to monitor deterioration of language abilities over time. Figure 3.2 shows that the Moderate-severe group perform nearly at floor for all subtests except semantic association and sentence comprehension II. This is critical for the transdiagnostic validity of the MLSE because language impairments in PPA deteriorate with disease progression (Weintraub et al., 1990). For example, future research might aim to compare language impairments in PSA with early- vs. late-stage PPA.

In conclusion, we have shown that the MLSE is hypersensitive to the presence of a language impairment in PSA. A potential drawback of this is an over-estimation of the severity of language impairments, even in supposedly mild forms of PSA. However, the MLSE is still able to differentiate milder cases from cases with moderate to severe impairments. Thus, the next stage in the development of the MLSE is to explore the validity of the MLSE in PPA and PSA concurrently, and to relate performance on the MLSE to neural correlates in both PSA and PPA. A key benefit for future research to capitalise on is using the MLSE as a platform for direct, systematic comparisons of PSA and PPA. As the first assessment tool to be developed for aphasia across aetiologies, the MLSE represents a unique opportunity to conduct further comparisons of these forms of aphasia.
Chapter 4 - Exploring transdiagnostic brain-behaviour relationships in aphasia using the Mini-linguistic State Examination

Ruth U. Ingram¹, Ajay D. Halai², Nikil Patel⁴, Katie Peterson³, Gorana Pobric¹, Karalyn Patterson²,³, Peter Garrard⁴ & Matthew A. Lambon Ralph²

¹Division of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, UK
²MRC Cognition & Brain Sciences Unit, University of Cambridge, UK
³Department of Clinical Neurosciences, University of Cambridge, UK
⁴Centre for Clinical Neuroscience, St George's University of London, UK

Statement of contribution
Ruth Ingram analysed all the behavioural and neuroimaging data presented in this chapter (including pre-processing), and additionally collected the behavioural data on the primary progressive aphasia cohort in Manchester. Blanca De Dios Perez collected the behavioural data on the chronic post-stroke aphasia cohort. Nikil Patel and Katie Peterson collected the behavioural data on the healthy controls, and primary progressive aphasia and movement disorders cohorts in London and Cambridge. Matt Lambon Ralph, Ajay Halai, and Gorana Pobric provided guidance and support throughout the analysis, and feedback on drafts of the write up.
Abstract

Acquired impairment in language comprehension or production, known as aphasia, can arise due to multiple aetiologies of brain injury, such as stroke or neurodegenerative disease. Conducting comparisons of these forms of aphasia should elucidate the brain-behaviour relationships that are fundamental to aphasia regardless of the nature of brain damage. Despite this, there are few systematic comparisons of aphasia across aetiologies. Therefore, this study directly compared linguistic and neural abnormalities across neurodegenerative and stroke aetiologies. Crucially, we compared language impairments on the same transdiagnostic assessment tool, situated these impairments in the same, shared multidimensional space, and analysed neural abnormality across aetiologies in the same voxel-based morphometry analysis. Specifically, we compared language impairments in post-stroke aphasia (PSA), progressive non-fluent aphasia (PNFA), logopenic progressive aphasia (LPA), semantic dementia (SD), corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP) using a novel clinical assessment tool, the Mini-Linguistic State Examination (MLSE). Principal component analysis was used to extract shared dimensions of variance in language impairments across these patient groups. We quantified neural abnormality (relative to healthy controls) using the same automated detection procedure in all patients. Thus, we were able to enter all patients into the same voxel-based correlational methodology analysis, to isolate transdiagnostic neural correlates of the shared behavioural dimensions. Overall, the PSP and CBS groups showed milder linguistic deficits, and abnormality restricted to more midbrain and subcortical regions, compared to the other groups. The PSA group showed the most extensive region of abnormal neural tissue and corresponding impairments in almost all MLSE subtests. PNFA, LPA and SD showed the expected pattern of linguistic deficits and patterns of neural abnormality. The principal component analysis revealed that language impairments across these aetiologies can be described in terms of variation in three key dimensions: phonology, semantics and syntax. Further, these dimensions were related to unique neural correlates in: left hemisphere superior and posterior temporal lobe, bilateral anterior temporal lobes, and left hemisphere subcortical regions and periventricular white matter, respectively. Thus, by comparing linguistic impairments on the same test battery, in the same multidimensional space, and using the same neuroimaging approach, we were able to extract brain-behaviour relationships that underly variance in aphasia across aetiologies.
Introduction

Aphasia is an impairment of the ability to comprehend and formulate language following acquired brain damage, which manifests as difficulties across multiple modalities of language use (e.g., reading, auditory comprehension, expressive language) (Rosenbek et al., 1989). Aphasia can arise due to multiple aetiologies of brain injury. Left-hemisphere stroke is a common cause of aphasia (Berthier, 2005), which we will refer to as post-stroke aphasia (PSA). Primary progressive aphasia (PPA) is the term given to aphasia which is the most salient symptom of neurodegenerative disease (Mesulam, 2001a). However, language impairments can also be a concomitant symptom in other neurodegenerative disorders, including progressive supranuclear palsy (PSP) (Höglinger et al., 2017) and corticobasal syndrome (CBS) (Frattali et al., 2000). Conducting comparisons of these forms of language impairment should elucidate the brain-behaviour relationships that are fundamental to aphasia regardless of the nature of brain damage. Despite this, there are few systematic comparisons of aphasia across aetiologies. Therefore, this study involved a large-scale comparison of aphasic profiles and neural correlates of impairment transdiagnostically across these stroke-related and neurodegenerative aetiologies.

Previous comparisons of aphasia across aetiologies have often been limited in the selection of behavioural presentations included. For example, comparing only semantically-impaired aphasics (Jefferies et al., 2008, Jefferies and Lambon Ralph, 2006), only ‘fluent’ aphasics (Silveri et al., 2019, Ogar et al., 2011) or only ‘non-fluent’ aphasics (Patterson et al., 2006a). In addition, rarely do these comparisons include aphasics with ‘mixed’ diagnoses (though Silveri et al. (2019) included logopenic/mixed PPA as their fluent PPA cases). Alternatively, comparisons between PSA and PPA often focus on a single aspects of language/cognition, such as naming (Budd et al., 2010), writing (Faria et al., 2013), syntactic processing (Thompson et al., 2013), or use of verbs and nouns (Thompson et al., 2012). In summary, there is a lack of studies which include the full phenotypic range of behavioural presentations of PSA and PPA and assess a broad range of linguistic/cognitive symptoms simultaneously.

Comparisons of PSA and PPA which consider structural brain abnormalities are even scarcer. Ogar et al. (2011) found that semantic dementia (SD) and Wernicke’s aphasia have distinct behavioural profiles, with Wernicke’s aphasia being more impaired than SD on spontaneous speech and sentence comprehension measures. These behavioural differences were related to neural damage; the region of greatest neural damage in SD was established using voxel-based morphometry, whilst lesions were manually demarcated by hand for Wernicke’s aphasia. These separate neuroimaging analyses showed differences in the location of neural damage (left anterior temporal lobe in SD and left posterior middle temporal gyrus in Wernicke’s aphasia). The authors conclude that the behavioural differences could be due to the different locations of damage. However, the difference in aetiology cannot be ruled out in this case since the aetiologies and locations both differed.
between the groups. Furthermore, the extent of damage cannot be compared directly across aetiologies since the brain scans were not part of the same analysis procedure.

Budd et al. (2010) compared naming errors in PPA and acute PSA to establish whether there were differences due to the aetiology or the location of the brain damage. Acute PSA were categorised based on the location of their stroke (middle cerebral artery (MCA) stroke either anterior, posterior or mixed, posterior cerebral artery, anterior cerebral artery, or subcortical). PPA were classified into a subtype (semantic variant PPA, non-fluent variant PPA, or logopenic variant PPA) using clinical and imaging criteria (Gorno-Tempini et al., 2011). The subtypes of PPA were matched with an acute PSA subtype based on the region of damage: anterior perisylvian cortex for nfvPPA with anterior MCA stroke, posterior perisylvian cortex for lvPPA with posterior MCA stroke, and anterior/inferior temporal cortex for svPPA with subcortical stroke. When naming performance was collapsed across location of damage, there were no significant differences between aetiologies. However, when naming performance was collapsed across aetiology, there were significantly fewer circumlocutions in the anterior MCA/anterior perisylvian cortex group than the other locations of damage. Despite, as noted by the authors, the location of brain damage being coarsely defined, this study highlights the utility of exploring aphasia transdiagnostically in order to investigate the fundamental aspects of aphasia which do not vary with aetiology (e.g., the neural correlates of language impairments).

There are key differences in the location and nature of brain injuries that cause PSA and PPA (Grossman and Irwin, 2018). Firstly, PSA is usually caused by unilateral stroke in the left hemisphere (Berthier, 2005), whilst neurodegeneration in PPA can be bilateral (though usually left-dominant) (i.e., in semantic dementia (Acosta-Cabronero et al., 2011) and late-stage logopenic progressive aphasia (Rohrer et al., 2013)). Secondly, the location of brain injury in PSA is dictated by the vascular territory of the occluded artery (Hillis, 2007), whilst neurodegeneration in PPA is thought to follow a pathology-specific spreading pattern (Leyton et al., 2016a). Consequently, the linguistic/cognitive deficits in PSA are somewhat restricted by the brain areas which are most vulnerable to stroke due to the nature of their blood supply (e.g., the temporal lobe is supplied by two arteries (Kiernan, 2012) and therefore may be less vulnerable to stroke). In contrast, the evolution of linguistic/cognitive deficits over time in PPA reflects the spread of atrophy to a diverse range of brain areas. Possible factors influencing which areas become affected could include the mechanism of pathological spread (e.g., neuron-to-neuron transmission (Jucker and Walker, 2013)), the epicentre from which pathology spreads (Seeley et al., 2009), and the selective vulnerability of entire brain networks (Warren et al., 2012).

Thirdly, the relative effect of stroke and neurodegeneration on white matter and subcortical structures is still being explored. Lack of blood supply in stroke indiscriminately affects grey and white matter, with deep white matter being especially vulnerable to ischaemia due to having less microvasculature (Nasrabady et al., 2018, Nonaka et al., 2003), and white
matter injury potentially accounting for up to half of the infarct volume (Ho et al., 2005). Subcortical infarcts may also account for a large majority (85%) of ischaemic strokes (Wessels et al., 2006). In contrast, neurodegenerative disease is principally studied in the context of grey matter degeneration (Romito-DiGiacomo et al., 2007). For example, degeneration of layers II (Gómez-Isla et al., 1996) and IV (Hyman et al., 1984) in Alzheimer’s disease, and layer II in fronto-temporal lobar degeneration (Mackenzie et al., 2006). However, more recent research has focused on white matter integrity in neurodegeneration (Seeley et al., 2009), in the context of white matter-dependent network-level changes in neurodegenerative disease, including PPA (Galuntucci et al., 2011, Mahoney et al., 2013, Schwindt et al., 2013) and Alzheimer’s disease (Nasrabady et al., 2018). In summary, PSA and PPA represent different but overlapping windows into the core underlying nature of aphasia. Employing a transdiagnostic approach to study aphasia across these two aetiologies could capitalise on these different perspectives.

A major benefit of transdiagnostic approaches is that they directly situate the impairments of two conditions in the same ‘space’, to be compared quantitatively. In Alzheimer’s disease, overlapping symptoms and atrophy patterns have been found in typical Alzheimer’s disease, posterior cortical atrophy and logopenic progressive aphasia (Ridgway et al., 2012, Migliaccio et al., 2009, Crutch et al., 2012b). In fronto-temporal lobar degeneration, Murley et al. (2019) compared the heterogeneous clinical presentations of behavioural variant fronto-temporal dementia, non-fluent variant PPA, semantic variant PPA, logopenic variant PPA, progressive supranuclear palsy, and corticobasal syndrome. Using principal component analysis (PCA) and multivariate source-based morphometry, they found shared symptom dimensions and co-varying regions of brain atrophy. In PSA, a robust picture of the multidimensional phenotypic space is emerging using PCA (Mirmans et al., 2015a, Lacey et al., 2017); patients with PSA appear to vary in terms of phonology, semantics, and speech fluency (Butler et al., 2014, Halai et al., 2017, Halai et al., 2018a). These principal dimensions of variance have been related to unique neural correlates in the left hemisphere using voxel-based correlational methodology (VBCM) (Halai et al., 2017); the phonology dimension is associated with lesions in left perisylvian language regions, the semantic dimension is associated with the left anterior temporal lobe, and the speech fluency dimension is linked with lesions in the left prefrontal lobe.

Taken together, these studies suggest that a multidimensional approach using PCA and VBCM could elucidate the brain-behaviour relationships of aphasia across neurodegenerative and stroke-related aetiologies, which has not been done before. Therefore, in the current study we leveraged the increased information available by including multiple aetiologies to explore the full phenotypic spectra of behavioural and neural variation in aphasia. Specifically, we explored linguistic impairments in PSA, PPA, corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP). PSP and CBS are classified as movement disorders, but a growing body of research highlights the significance of often
under-recognised language impairments in these conditions (Peterson et al., 2019). The language impairments PSA, PPA, CBS and PSP have not been simultaneously compared using the same test battery. We compared these groups using a novel clinical assessment tool, the Mini-Linguistic State Examination (MLSE), because this tool is being developed to have equivalent validity for assessing language impairments caused by stroke or neurodegeneration. MLSE performance for all groups was entered into a principal component analysis to reveal the shared dimensions of variance in aphasia across aetiologies. As noted above, comparisons of structural brain changes in PSA and PPA tend to use different methods to estimate the extent of lesioned/atrophied neural tissue. Therefore, to situate the brain-behaviour relationships in aphasia across aetiologies in the same ‘space’ we analysed all structural brain scans in the same voxel-based correlational methodology analysis. We applied an automated abnormality detection tool (Seghier et al., 2008) to all brains to define regions of abnormality relative to controls. The principal components reflecting MLSE performance were entered as covariates in the voxel-based correlational methodology to uncover the unique neural correlates of language impairments across aetiologies.

Methods

Participants
A total of one hundred and five people participated in this study (for breakdown per subgroup, see below), prospectively recruited from three centres across the United Kingdom. All participants completed the Mini-Linguistic State Examination and had a structural magnetic resonance imaging (MRI) scan of their brain. Recruitment details of each cohort are outlined below. All participants gave informed consent under approval from respective local ethics committees; University of (North West Multi-Centre Research Ethics Committee, UK); University of Cambridge (Cambridge Central Research Ethics Committee); St George’s University of London (London-Chelsea Research Ethics Committee).

Healthy controls
Twenty people were recruited as matched healthy control (HC) participants from St. George’s University of London and University of Cambridge. Healthy controls were excluded from participation if they reported a history of neurological, psychological, or speech, language, and/or learning deficits. All healthy controls were native English speakers with normal or corrected to normal hearing and vision.

Patients
Primary progressive aphasia
Forty-two people with a diagnosis of PPA were recruited prospectively across three centres for this study. From the Manchester centre, six people were recruited from specialist neurology clinics, old age psychiatry clinics, memory clinics, and community groups in the North West of England. From the St. George’s University of London centre, 18 people were
recruited from a specialist neurology clinic. Seventeen people were recruited from the Cambridge centre. Diagnosis of PPA was based on current clinical criteria (Gorno-Tempini et al., 2011) after comprehensive neurological, neuropsychological and speech and language assessment by a trained neurologist. PPA was diagnosed when language impairment was the primary symptom of a neurodegenerative disorder for at least two years (Mesulam, 2001a). There remains some contention surrounding the terminology used to label the three most common subtypes of PPA. We will refer to them as progressive non-fluent aphasia (PNFA) (Neary et al., 1998), semantic dementia (SD) (Snowden et al., 1989), and logopenic progressive aphasia (LPA). This cohort included 15 people with LPA, 13 people with PNFA, and 14 people with SD. Exclusion criteria included other causes of aphasia (e.g., non-neurodegenerative pathology), non-native English speakers, and any other neurological or major psychiatric illness.

Post-stroke aphasia
Twenty-one people with chronic post-stroke aphasia (PSA) were recruited prospectively from community groups and speech and language therapy services in the North West of England. Participants were included if they reported a single left hemisphere stroke at least 12 months prior to assessment and were native English speakers. People with chronic post-stroke aphasia were recruited at all levels of severity. This cohort included a heterogeneous collection of aphasic phenotypes, classified using the Boston Diagnostic Aphasia Examination (Kaplan, 1983) as: anomic aphasia (N = 10), Broca’s aphasia (N = 3), conduction aphasia (N = 3), global aphasia (N = 2), mixed non-fluent aphasia (N = 2) and transcortical motor aphasia (N = 1).

Corticobasal syndrome and progressive supranuclear palsy
Six people with corticobasal syndrome and nine people with progressive supranuclear palsy were recruited through the University of Cambridge from research cohorts and a specialist clinic. Five people with corticobasal syndrome and two people with progressive supranuclear palsy were recruited from St. George’s University of London specialist neurology clinic, bringing the totals to 11 for both groups. Diagnosis of corticobasal syndrome (CBS) was made using established criteria (Armstrong et al., 2013). We note that this diagnosis was based on clinical presentation not pathology and therefore the label of corticobasal degeneration (which is limited to pathology) was not applied. Established criteria were also used to diagnose progressive supranuclear palsy (PSP) (Höglinger et al., 2017).

Mini-linguistic state examination
The Mini-Linguistic State Examination is a language assessment battery comprising tests to measure the major linguistic domains commonly affected in primary progressive aphasia (Gorno-Tempini et al., 2011). The MLSE was designed to be a brief yet comprehensive test for use in clinical settings, and as such short completion time was prioritised (the MLSE can be administered in 20 minutes on average, with 5 further minutes for calculating the scoring).
Therefore, a small selection of items (difficult items or of graded difficulty) are included for each subtest (minimum of 3, maximum of 6).

The 13 subtests are intended to capture variation in fundamental aspects of language production and comprehension, namely phonology, semantics, syntax, and motor speech. To assess these capacities, the MLSE subtests are: picture naming; repetition of syllables, words, non-words and sentences; sentence comprehension (I and II); semantic memory (spoken word-picture matching and semantic association); reading words with irregular sound-spelling associations (e.g., 'mauve') and reading non-words; procedural writing ('explain to someone how to brush their teeth'); and spoken picture description (beach scene). Sentence comprehension was assessed through two subtests; sentence comprehension I involves participants listening to a sentence with complex syntactic structure, followed by a question relating to the sentence (e.g., ‘Joe was treated by Mary. Who was the doctor?’) that is read by the experimenter. The participant answers verbally. Sentence comprehension II also involves a syntactically-complex sentence spoken by the experimenter, which the participant must match to a picture illustrating that sentence. The target picture is accompanied by three distractors which contain the same action/actors in different confirmations. Word repetition and spoken word-picture matching mimic each half of the ‘repeat and point’ test (Hodges et al., 2008), wherein the participant repeats a single word after the experimenter, then points to the picture corresponding to that word (from a selection including five distractors). The sentence repetition test was of graded difficulty, increasing from six to 15 words in length, and also contained sentences with an ‘unexpected’ conclusion (e.g., ‘the astronomer gazed at the flowers’).

Data analysis

For the purpose of all analyses, patients were split into the following subgroups: healthy controls (HC), PSA, PNFA, LPA, SD, CBS, and PSP. Accuracy data on the MSLE subtests were converted to percentages. There were no missing data. Statistical analyses were conducted in SPSS 25 (IBM statistics). Demographic variables were compared across subgroups using one-way Analysis of Variance (ANOVA) (age, education, symptom duration, lesion size) and chi-squared tests (sex distribution).

One-way Multivariate ANOVA (MANOVA) was used to assess differences in accuracy on the MLSE subtests across subgroups. One-way MANOVA tests for differences across levels of the independent variable (subgroups) in terms of a linear composite of their scores across all dependent variables (MLSE subtests). Univariate ANOVAs (with Bonferroni correction for multiple comparisons) were used to test a significant one-way MANOVA result. If there was a significant univariate ANOVA result, post-hoc pairwise comparisons were conducted to assess differences at the level of error types/MLSE subtests per subgroup.

In summary, to explore differences in the subgroup profiles of accuracy across MLSE subtests, we conducted a 7 x 13 MANOVA (7 subgroups, 13 MLSE subtests). Following this,
we conducted seven univariate (1 x 13) ANOVAs to explore which subgroups showed a significant difference across performance on the MSLE subtests. When the assumption of homogeneity of variance was violated (e.g., significant result of Mauchly’s test for sphericity or Levene’s test for homogeneity of variance) appropriate corrections were applied to the results of (M)ANOVAs and t-tests. For each significant ANOVA result, we conducted 78 pairwise multiple comparisons t-tests across MLSE subtests (13 subtests x 7 groups). Bonferroni correction for multiple comparisons was applied, with significant differences being considered at p < .05 after correction.

Principal component analysis

The MLSE data were entered into PCA (conducted in SPSS 25), with correlation decomposition, using varimax rotation to aid cognitive interpretation of the extracted dimensions. We inspected the pattern of loadings onto each factor to help with interpretation. Factor scores were generated per participant using regression. Selecting the number of components for PCA to extract is a non-trivial process. A simple cut-off of eigenvalue >1 can be used (Kaiser, 1958). However, we used a data-driven method utilizing a freely-available toolbox in MATLAB (R2017a) (Ballabio, 2015). This toolbox uses cross-validation with Venetian blind sampling. First, the toolbox builds a PCA model on the training set (z-score scaling to all variables) and is used to predict the variables in the hold out set. One variable is removed at a time and these held-out data are predicted using linear regression from the PCA model. The residuals of this reconstruction are calculated as root mean squared error in cross-validation and analysed as a function of the number of components (total N-1, where N = number of test variables). We repeated this step 1000 times where, on each occasion, the order of the patients was shuffled to avoid biases in the Venetian blinds sampling approach. The final number of components to extract was chosen based on the solution with the lowest root mean squared error.

Neuroimaging

Acquisition

University of Manchester

Post-stroke aphasia

High resolution structural T1-weighted Magnetic Resonance Imaging (MRI) scans were acquired on a 3.0 Tesla Philips Achieva scanner (Philips Healthcare, Best, The Netherlands). A T1-weighted inversion recovery sequence with 3D acquisition was employed, with the following parameters: TR (repetition time) = 9.0 ms, TE (echo time) = 3.9 ms, flip angle = 8°, 150 contiguous slices, slice thickness = 1 mm, acquired voxel size 1.0 x 1.0 x 1.0 mm³, matrix size 256 x 256, FOV = 256 mm, TI (inversion time) = 1150 ms, SENSE acceleration factor 2.5.
High resolution structural T1-weighted MRI scans were acquired on a 3.0 Tesla Philips Achieva scanner (Philips Healthcare, Best, The Netherlands). A T1-weighted inversion recovery sequence with 3D acquisition was employed, with the following parameters: TR (repetition time) = 6.6 ms, TE (echo time) = 2.9 ms, flip angle = 8°, 182 contiguous slices, slice thickness = 1.1 mm, acquired voxel size 1.1 × 1.1 × 1.1 mm³, matrix size 256 × 256, FOV = 200.2 mm, TI (inversion time) = 1150 ms, SENSE acceleration factor 2.5.

University of Cambridge

High resolution structural T1-weighted MRI scans were acquired on a 3.0 Tesla SIEMENS Prisma Fit scanner (SIEMENS Medical Solutions, Erlangen, Germany). A T1-weighted inversion recovery sequence with 3D acquisition was employed, with the following parameters: TR (repetition time) = 2000 ms, TE (echo time) = 2.9 ms, flip angle = 8°, 154 contiguous slices, slice thickness = 1.1 mm, acquired voxel size 1.0 × 1.0 × 1.0 mm³, matrix size 256 × 256, FOV = 281.6 mm, TI (inversion time) = 850 ms, GRAPPA acceleration factor = 2.

St. George’s University of London

High resolution structural T1-weighted MRI scans were acquired on a 3.0 Tesla Philips Achieva scanner (Philips Healthcare, Best, The Netherlands). A T1-weighted inversion recovery sequence with 3D acquisition was employed, with the following parameters: TR (repetition time) = 6.7 ms, TE (echo time) = 3.0 ms, flip angle = 8°, 182 contiguous slices, slice thickness = 1.1 mm, acquired voxel size 1.1 × 1.1 × 1.1 mm³, matrix size 256 × 256, FOV = 200.2 mm, TI (inversion time) = 1150 ms, SENSE acceleration factor 2.5.

Analysis

Pre-processing

T1 structural MRI scans were pre-processed with Statistical Parametric Mapping software (SPM12: Wellcome Trust Centre for Neuroimaging), running in MATLAB (R2017a). Pre-processing was conducted using an SPM toolbox called Automated Lesion Identification (version 3.0) (Seghier et al., 2008), which combines segmentation, bias correction and spatial normalisation through the inversion of a single unified model (Ashburner and Friston, 2005). Although considered as a ‘lesion’ identification method, the technique fundamentally identifies areas of abnormal tissue class relative to controls, and so detects abnormal grey matter, white matter, and cerebral spinal fluid (CSF) space. Therefore, this automated abnormality detection procedure was deemed suitable for exploratory investigation of the correlations between behavioural variation and neural abnormality across aetiologies of aphasia (rather than the correlates for lesioned/atrophied brain specifically).

In brief, the unified model combines tissue class (with an additional tissue class for abnormal voxels), intensity bias and non-linear warping into the same probabilistic models that are assumed to generate subject-specific images. All images were normalised into standard
Montreal Neurological Institute (MNI) space. The regions of neural abnormality of each patient were automatically identified using an outlier detection algorithm, compared to healthy controls, based on fuzzy clustering. Separate groups of age- and education-matched healthy controls were used for this pre-processing step for the PSA and non-PSA cohorts; healthy controls for the PPA, CBS, and PSP groups were the same as reported in the behavioural analysis, whilst specific healthy controls were recruited to be matched to the PSA cohort (see Table 4.1). The default parameters of the automated ‘lesion’ detection procedure were used to create a binary abnormality image. Images were then smoothed with an 8 mm full-width-half-maximum (FWHM) Gaussian kernel and used in the analyses described below.

Subgroup significant abnormality maps
Average abnormality maps per subgroup were generated to confirm that the automated procedure was capable of detecting neural abnormality in brains with neurodegeneration (as opposed to stroke lesions). To do this, the regions of significant abnormality for each subgroup were compared to expected atrophy patterns reported in the literature. The abnormality maps represent how deviant the T1 signal within a voxel was compared to controls. Each map was smoothed and averaged based on subgroup membership. The same smoothed images per patient were entered into a whole-brain analysis using one-sample t-tests. The resultant statistical maps showed regions of significant abnormality. The outline of the region of significant abnormality was then overlayed onto the mean abnormality map to give a sense of which regions that were abnormal in most patients were significantly abnormal (Figure 4.3).

A combined atlas for grey and white matter was used to label locations of significant neural abnormality (Halai et al., in press). This atlas combined the Harvard-Oxford (Desikan et al., 2006) and Johns Hopkins University (Mori et al., 2005) atlases to cover cortical and subcortical grey and white matter, but note the combined atlas does not include cerebellum or third ventricle.

Voxel-based correlational methodology
Continuous signal intensity values from ‘fuzzy lesion’ images (i.e., images normalised and scaled against the same controls as above) of the whole brain were correlated with PCA factor scores and other covariates (see below) using a voxel-based correlational methodology (VBCM) (Tyler et al., 2005) (conducted in SPM12). This is a variant of voxel-lesion symptom mapping (VSLM) (Bates et al., 2003). Unlike VLSM, VBCM does not require a binary classification of the intact/abnormal neural tissue because both the behaviour and tissue concentration measures are treated as continuous variables.

All participants’ factor scores from the PCA were entered into the VBCM to investigate how variation in tissue abnormality corresponded to the unique components reflecting phonology, semantics, and syntax. A study-specific explicit brain mask was applied. In order to ensure
that the results were not merely attributable to the degree of overall abnormality, we calculated each participant's 'lesion' volume as the sum of abnormal voxels identified by the automated identification method (Seghier et al., 2008) and this was entered as a covariate in the VBCM. Age, symptom duration and a categorical variable for scanning protocol were also all included as covariates in the VBCM.

Results

Demographic details

Demographic details are summarised in Table 4.1. There was no significant difference between subgroups in sex distribution ($\chi^2(6) = 1.92, p = .927$), age ($F(6,40.5) = 2.19, p = .064$ (Welch)), or lesion size ($F(5,34.2) = 2.46, p = .052$ (Welch)). The main effect of Education was significant ($F(6,104) = 3.50, p = .004$), but all pairwise comparisons to explore this effect were non-significant after correction for multiple comparisons. There was a significant difference between patient subgroups in symptom duration ($F(5,84) = 6.50, p < .001$) with patients with PSA having significantly longer disease duration than LPA ($p < .001$) and CBS ($p < .001$).

Table 4.1 – Demographic details of all participants.

<table>
<thead>
<tr>
<th>Group</th>
<th>N (F)</th>
<th>Age</th>
<th>Education</th>
<th>Symptom duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC(1)</td>
<td>20 (7)</td>
<td>63.5 (6.4)</td>
<td>15.5 (3.0)</td>
<td>-</td>
</tr>
<tr>
<td>HC(2)</td>
<td>22 (10)</td>
<td>69.1 (6.0)</td>
<td>13.0 (2.6)</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>14 (5)</td>
<td>67.1 (5.3)</td>
<td>14.3 (2.5)</td>
<td>5.7 (4.4)</td>
</tr>
<tr>
<td>PNFA</td>
<td>13 (7)</td>
<td>69.8 (6.1)</td>
<td>12.7 (2.3)</td>
<td>3.7 (2.9)</td>
</tr>
<tr>
<td>LPA</td>
<td>15 (5)</td>
<td>70.2 (7.3)</td>
<td>14.4 (3.3)</td>
<td>3.1 (2.0)*</td>
</tr>
<tr>
<td>CBS</td>
<td>11 (5)</td>
<td>70.7 (9.6)</td>
<td>12.9 (3.3)</td>
<td>2.8 (1.8)*</td>
</tr>
<tr>
<td>PSP</td>
<td>11 (5)</td>
<td>67.5 (4.8)</td>
<td>11.9 (1.9)</td>
<td>4.0 (2.3)</td>
</tr>
<tr>
<td>PSA</td>
<td>21 (9)</td>
<td>64.8 (11.3)</td>
<td>12.4 (2.7)</td>
<td>8.5 (4.8)</td>
</tr>
</tbody>
</table>

Years of age, education and symptom duration are expressed as mean (standard deviation). Abbreviations: HC(1) – healthy controls for MLSE data and neurodegenerative neural abnormality detection (see Methods); HC(2) – healthy controls for stroke neural abnormality detection (see Methods); SD – semantic dementia; PNFA – progressive non-fluent aphasia; LPA – logopenic progressive aphasia; CBS – corticobasal syndrome; PSP – progressive supranuclear palsy; PSA – post-stroke aphasia. * significantly different from post-stroke aphasia ($p < .05$ Bonferroni corrected).

Profiles of impairment on the MLSE

The profile of scores over the MLSE per subgroup is shown in Figure 4.1 (statistically significant differences between MLSE subtests and between subgroups are described in the text as the figures became crowded with markers indicating the multitude of post-hoc t-test results). Differences between MLSE scores across subgroups were explored using one-way multivariate analysis of variance (MANOVA). The accuracy on each of the 13 MSLE subtests was compared for all seven subgroups. Data are expressed as mean ± standard deviation. The data were found to be suitable for MANOVA; Preliminary assumption
checking revealed that several variables were not normally distributed, as assessed by Shapiro-Wilk test (p < .05) and so analyses on these variables are reported with appropriate correction. There was no multicollinearity, as assessed by Pearson correlations (highest correlation was between Irregular word reading and Naming (r = 0.759, p < .001)). There was a single multivariate outlier (PNFA), as assessed by Mahalanobis distance (p > .001). Removing this case did not substantially alter the result of the MANOVA so the results with this case retained are shown. There was homogeneity of variance-covariances matrices, as assessed by Box’s M test of equality of covariance matrices (p = .049). The assumption of homogeneity of variances was met for all MLSE subtests except Sentence Comprehension I (p = .055), Sentence Repetition (p = .095) and Picture Description (p = .331). For these three subtests, follow-up univariate ANOVAs and post-hoc tests were corrected for violating this assumption.

**Post-stroke aphasia**
Participants with PSA were significantly impaired on all MLSE subtests relative to HC apart from Semantic Association (p = .211). Performance on naming, non-word repetition, non-word reading, and sentence repetition were additionally significantly worse than PSP. Performance on non-word reading, word repetition, and non-word repetition were also significantly worse than SD, whilst semantic association performance was better than SD. Within PSA, there was a significant main effect of MLSE subtest (F(6.14,122.81) = 19.459, p < .001, partial $\eta^2 = .493$ (Greenhouse-Geisser)), with semantic association performance being significantly better than naming, repetition (non-words and sentences), reading (irregular words and non-words), sentence comprehension I (auditory working memory demands), writing, and picture description. Additionally, non-word repetition was also significantly impaired compared to spoken word-picture matching. These results broadly show that the PSA group were more impaired on tests of repetition than semantic knowledge.

**Progressive non-fluent aphasia**
Compared to controls, participants with PNFA were significantly impaired on naming, syllabic repetition, non-word repetition, sentence comprehension I and II, sentence repetition, non-word reading, writing, and picture description. Sentence repetition was additionally worse than PSP. Although the univariate repeated-measures ANOVA for MLSE subtests was significant (F(4.9,58.9) = 4.15, p = .003, partial $\eta^2 = .257$ (Greenhouse-Geisser)), none of the pairwise comparisons survived correction for multiple comparisons.

**Logopenic progressive aphasia**
The LPA group performed significantly worse than controls on naming, non-word repetition, sentence comprehension I (auditory working memory demands), sentence repetition, writing, and picture description. Additionally, LPA patients performed worse at sentence repetition than PSP. These results highlight that the LPA showed worse performance on tests of sentence-level processing (sentence repetition, sentence comprehension, writing, picture description), than semantic knowledge. Although the univariate repeated-measures ANOVA
for MLSE subtests was significant ($F_{(12,168)} = 10.241, p < .001, \text{ partial } \eta^2 = .422$), none of the pairwise comparisons survived correction for multiple comparisons.
Figure 4.1 – MLSE performance per subgroup.

Coloured bars represent mean (± standard error) performance on each MLSE subtest for patient subgroups. The grey line indicates mean (± standard error) for healthy controls. Groups are colour coded: PSA = purple; PNFA = light blue; LPA = dark blue; SD = black; CBS = yellow; PSP = orange. Abbreviations: PSA = post-stroke aphasia; PNFA = progressive non-fluent aphasia; LPA = logopenic progressive aphasia; SD = semantic dementia; CBS = corticobasal syndrome; PSP = progressive supranuclear palsy; sWPM = spoken word-to-picture matching.
**Semantic dementia**

Participants with semantic dementia performed significantly worse at both naming and spoken word-picture matching compared to controls and compared to participants with CBS and PSP. Semantic Association was significantly impaired relative to controls but also relative to PSA and PSP. Irregular word reading, sentence repetition, sentence comprehension I (auditory working memory demands), writing, and picture description were also significantly impaired compared to controls. Non-word reading and repetition (words, non-words) performance was better in SD than PSA. The univariate repeated-measures ANOVA revealed significantly different performance across the MLSE subtests ($F_{(12,156)} = 19.075$, $p = .001$, partial $\eta^2 = .595$). Specifically, performance in SD was significantly worse for naming compared to repetition (syllables, words, non-words) and sentence comprehension II (with picture stimuli). Furthermore, picture description performance was also worse in SD compared to repetition of syllables and words. Overall, these results highlight the specific impairment in semantic memory experience in SD, with relative sparing of ability to repeat.

**Corticobasal syndrome**

Participants with corticobasal syndrome did not perform significantly worse than controls on any MLSE subtest. CBS patients performed significantly better than SD on naming and spoken word-picture matching. The performance of CBS patients did not vary significantly across the MLSE subtests ($F_{(12,120)} = 0.950$, $p = .500$, partial $\eta^2 = .087$). However, we also conducted targeted pairwise t-tests (CBS vs. HC) guided by a priori hypotheses regarding known language impairments in CBS (Santos-Santos et al., 2016), speech production (Özsancak et al., 2000), single-word and sentence repetition (Burrell et al., 2013), non-word reading (Graham et al., 2003) and writing (Gorno-Tempini et al., 2004)). Correcting for seven pairwise comparisons, CBS patients were significantly impaired relative to controls for sentence comprehension I and writing.

**Progressive supranuclear palsy**

Participants with PSP did not perform significantly differently to controls on any MLSE subtest. Naming, spoken word-picture matching, semantic association performance was better in PSP than SD. Naming, non-word repetition, non-word reading, and sentence repetition performance was also better in PSP than PSA. PSP patients performed significantly better at sentence repetition than PNFA, LPA and PSA. Within PSP, although the univariate repeated-measures ANOVA showed a significant main effect of MLSE subtest ($F_{(12,120)} = 5.312$, $p < .001$, partial $\eta^2 = .347$), none of the pairwise comparisons were significant after correction for multiple comparisons. However, as with CBS, we conducted targeted pairwise t-tests (PSP vs. HC) guided by a priori hypotheses concerning reported language problems in PSP (speech production (Sakai et al., 2002), semantic association (Catricalà et al., 2019), repetition of single words (Burrell et al., 2018), oral reading (Podoll et al., 1991), and writing (Sitek et al., 2015)). Correcting for six pairwise comparisons, PSP
were impaired relative to controls on reading (words and non-words), writing, and picture description (as a measure of speech production).

**Shared principal components of aphasia**

The PCA was robust (Kaiser Meyer-Olkin = 0.876) and produced a 3-factor rotated solution which accounted for 69.7% of variance in patients' performance (F1 = 26.5%, F2 = 22.9%, F3 = 20.3%). The factor loadings of each behavioural assessment onto the extracted components are shown in Table 4.2. Before proposing cognitive interpretations for the extracted factors, it is useful to highlight that PCA decomposes tasks into their more fundamental processes and groups them together according to shared variance across the study population. For example, spoken picture naming is a task which is dependent on multiple systems (e.g., vision, semantic knowledge, phonological processing, articulation), and thus impairment in any of these systems can result in poor performance on this task. PCA decomposes the variance across all tests and is able to group them into factors which can reflect the underlying systems (e.g., semantic knowledge). This means that it is possible for a complex task, such as naming, to have split loadings across multiple principal components. These split loadings can be leveraged to aid the cognitive interpretation of the components.

Measures with strong loadings onto the first factor included all the tests of repetition, at the syllable, word, non-word, and sentence level. The strongest loading (0.843) onto this factor was word repetition. Tests of reading ability for non-words and irregular words also had strong loadings onto the first factor. Irregular word reading had split loadings across the first and second factors, whilst sentence repetition had split loadings across the first and third factors; accounting for variance in the following components (semantics and syntax, see below), irregular word reading and sentence repetition share the demands of producing phonetically-accurate speech. A key underlying system shared by the tests loading onto the first factor is phonological processing, hence we labelled this factor ‘Phonology’.

The second factor comprised strong loadings from tests tapping semantic knowledge. These included naming, spoken word-picture matching and semantic association. The strongest loading (0.820) was for spoken word-picture matching. There were also split loadings for irregular word reading (as above), and picture description which was split across the second and third factors. Irregular word reading relies on semantic knowledge to inform the exceptional sound-spelling relationships. Picture description also involves semantic knowledge since ability to recognise and name the items in the picture is required in order to describe it accurately. Coupled with the strong loadings from relatively more ‘pure’ tests of semantics (i.e., semantic association and word-picture matching), we labelled this factor ‘Semantics’.

The third factor comprised strong loadings from tests of sentence comprehension, with the sentence comprehension II test having the strongest loading on this factor (0.802). The
sentence comprehension II test involves selecting a picture which matches the sentence spoken by the experimenter. Other strong loadings onto this factor included writing, picture description and sentence repetition. Given that phonological processing and semantic knowledge have been accounted for in the previous two factors, the remaining variance in picture description and sentence repetition (which have split loadings) could reflect aspects of connected speech production or narrative level discourse. In the context of the strong loadings from sentence comprehension II (which does not require any verbal response), this third factor appears to reflect grammatical competence or awareness/use of correct syntax, which is captured in the context of production and/or comprehension of connected speech. We labelled this factor, ‘Syntax’.

Table 4.2 – Principal component analysis of MLSE performance across aetiologies.

<table>
<thead>
<tr>
<th>Test</th>
<th>Phonology</th>
<th>Semantics</th>
<th>Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllabic repetition</td>
<td>0.779</td>
<td>0.216</td>
<td>0.200</td>
</tr>
<tr>
<td>Word repetition</td>
<td>0.843</td>
<td>0.105</td>
<td>0.118</td>
</tr>
<tr>
<td>Non-word repetition</td>
<td>0.793</td>
<td>0.007</td>
<td>0.333</td>
</tr>
<tr>
<td>Sentence repetition</td>
<td>0.593</td>
<td>0.221</td>
<td>0.516</td>
</tr>
<tr>
<td>Non-word reading</td>
<td>0.691</td>
<td>0.199</td>
<td>0.434</td>
</tr>
<tr>
<td>Irregular word reading</td>
<td>0.507</td>
<td>0.673</td>
<td>0.057</td>
</tr>
<tr>
<td>Naming</td>
<td>0.336</td>
<td>0.813</td>
<td>0.169</td>
</tr>
<tr>
<td>Word-picture matching (spoken)</td>
<td>0.153</td>
<td>0.820</td>
<td>0.111</td>
</tr>
<tr>
<td>Semantic association</td>
<td>-0.244</td>
<td>0.740</td>
<td>0.263</td>
</tr>
<tr>
<td>Picture description</td>
<td>0.253</td>
<td>0.543</td>
<td>0.534</td>
</tr>
<tr>
<td>Writing</td>
<td>0.230</td>
<td>0.371</td>
<td>0.674</td>
</tr>
<tr>
<td>Sentence comprehension I</td>
<td>0.197</td>
<td>0.242</td>
<td>0.726</td>
</tr>
<tr>
<td>Sentence comprehension II</td>
<td>0.257</td>
<td>-0.023</td>
<td>0.802</td>
</tr>
</tbody>
</table>

This table shows the factor loadings of each test onto the extracted components. Factor loadings > .4 are shaded in grey, and factor loadings > .5 are additionally in bold.
| Sentence comprehension I | 0.197 | 0.242 | 0.726 |
| Sentence comprehension II | 0.257 | -0.023 | 0.802 |

The factor scores of all cases, colour coded by diagnostic subgroup, are shown in Figure 4.2. Panel A maps the phonology and semantics dimensions, Panel B maps the phonology and syntax dimensions, and Panel C maps the semantics and syntax dimensions. Plotting all cases’ factor scores on these three dimensions highlighted overlapping profiles in language impairments across these aetiologies of aphasia. Panels A and B show complete overlap between PNFA and PSA in terms of phonological impairments. Panels A and C show that semantic dementia occupy a relatively unique part of the space signifying spared phonological processing or syntax processing with selectively impaired semantics (though some CBS occupy the same space in Panel A). In Panels A, B and C, the movement disorders (CBS and PSP) groups tend to occupy the top right quadrant of the scatterplots, signifying that their language impairments are milder in these three dimensions that the other groups. However, in Panels B and C some cases of CBS and PSP do show lower scores on the syntax component. In Panel B, the region signifying poor syntax, but relatively preserved phonology is occupied mostly by cases of PNFA, LPA, and PSP. Overall, visualising behavioural variance in this way highlights the overlap in language impairments across many forms of aphasia.
Figure 4.2 – Shared multidimensional space of language impairments across aetiologies.

Factor scores of all patients were plotted along all pairs of components extracted by PCA. The origin is the mean of all patients. The factor scores are an expression of how many standard deviations a patient’s performance is from the group mean. Groups are colour coded: PSA = purple; PNFA = light blue; LPA = dark blue; SD = black; CBS = yellow; PSP = orange. Abbreviations: PSA = post-stroke aphasia; PNFA = progressive non-fluent aphasia; LPA = logopenic progressive aphasia; SD = semantic dementia; CBS = corticobasal syndrome; PSP = progressive supranuclear palsy.
**Subgroup atrophy maps**

The average abnormality maps per subgroup are presented in Figure 4.3 with an arbitrary scale of 15-25% abnormality relative to controls (purple to red scale). The regions of significant atrophy (see Methods) are outlined with a black contour line.

Abnormal voxels for PSA covered a large portion of the left-hemisphere, covering the insula, operculum cortex, planum temporale and polare, Heschl’s gyrus, posterior middle and superior temporal gyri, inferior frontal gyrus (pars opercularis and triangularis), superior and middle frontal gyri, posterior supramarginal gyrus, precentral gyrus, lateral ventricle, supplementary motor cortex and anterior cingulate gyrus. Abnormal voxels also covered subcortical regions, including the left thalamus, caudate, putamen and pallidum. Deep white matter tracts also contained abnormal voxels, such as the anterior thalamic radiations, and corticospinal tract. Abnormal voxels extended ventrally to cover the uncinate fasciculus, inferior frontal occipital fasciculus (IFOF), and inferior and superior longitudinal fasciculi (ILF/SLF). Significantly abnormal voxels were confined to the IFOF, SLF, ILF, uncinate fasciculus, anterior thalamic radiation, corticospinal tract, central and frontal operculum cortex, insula, putamen, caudate, inferior frontal gyrus (pars opercularis), and precentral gyrus.

In PNFA, abnormal voxels were found in a left-dominant pattern encompassing left posterior temporal cortex and underlying white matter. In the left hemisphere only, abnormal voxels extended across the inferior frontal occipital fasciculus and inferior longitudinal fasciculus, reaching the thalamus, and caudate. Abnormal voxels were found in the left posterior supramarginal gyrus, superior frontal gyrus and middle frontal gyrus. Voxels which were less abnormal relative to controls (purple spectrum) were found in left precentral gyrus. Bilaterally, both lateral ventricles, anterior thalamic radiations, and the anterior and posterior cingulate cortices had abnormal voxels. Bilateral but left-dominant insula and frontal operculum cortex also had abnormal voxels. Significantly abnormal voxels were restricted to bilateral anterior thalamic radiation, bilateral lateral ventricles, left caudate, left thalamus, left corticospinal tract, left IFOF, left SLF, and left ILF.

LPA displayed abnormal voxels in an asymmetrical, left dominant, pattern encompassing white matter passing through the posterior temporal lobe, regions surrounding the insula, and posterior temporal lobe structures. Specifically, posterior middle temporal gyrus, posterior superior temporal gyrus and posterior temporal fusiform cortex, with underlying white matter tracts (inferior and superior longitudinal fasciculi), were found to have abnormal voxels. Although less abnormal relative to controls (purple spectrum), the left dorsal angular gyrus also had abnormal voxels. Abnormal voxels in the left inferior longitudinal fasciculus extended ventrally to inferior frontal occipital fasciculus, planum polare, reaching the temporal pole in the left hemisphere. Abnormal voxels were found in the left insula, central operculum cortex, Heschl’s gyrus and planum temporale. In the right hemisphere, there were also abnormal voxels in the inferior frontal occipital fasciculus and superior longitudinal...
fasciculus, but to a lesser extent than the left hemisphere. Abnormal voxels were found in bilateral amygdala, but only left parahippocampal gyrus. Regions including and surrounding both lateral ventricles also had abnormal voxels, including anterior thalamic radiation, and corticospinal tracts. Significantly abnormal voxels were found in left planum polare, Heschl’s gyrus, left ILF, bilateral IFOF, bilateral SLF, forceps major, bilateral lateral ventricles, bilateral anterior thalamic radiation, and bilateral corticospinal tracts.

Abnormal voxels in SD were largely found in the bilateral anterior temporal lobes, but to a greater extent in the left hemisphere. Abnormal voxels in the left anterior temporal lobe encompassed the temporal pole, anterior parahippocampal gyrus and hippocampus, amygdala, and temporal fusiform cortex (anterior and posterior). The white matter pathways of the ventral language route in the left hemisphere, the inferior longitudinal fasciculus, inferior frontal occipital fasciculus and uncinate fasciculus, also contained abnormal voxels. This extended to the left hemisphere insula. Also, in the left hemisphere, the middle temporal gyrus and superior temporal gyrus had abnormal voxels in posterior and anterior regions. Although less abnormal relative to controls (purple spectrum colours), the left frontal pole and left postcentral gyrus also had abnormal voxels. In the right hemisphere, abnormal voxels were only found in the anterior temporal lobe (and to a lesser extent than left-hemisphere), including temporal pole, anterior parahippocampal gyrus, temporal fusiform cortex (anterior and posterior), inferior longitudinal fasciculus, hippocampus, amygdala, and inferior temporal gyrus (anterior and posterior). Significantly abnormal voxels were restricted in the right hemisphere to anterior and posterior inferior temporal gyrus, anterior and posterior temporal fusiform cortex, and anterior parahippocampal gyrus. In the left hemisphere, a larger regions of voxels were significantly abnormal, encompassing left temporal pole, anterior and posterior temporal fusiform cortex, anterior and posterior inferior temporal gyrus, anterior superior temporal gyrus, anterior and posterior middle temporal gyrus, anterior parahippocampal gyrus, hippocampus, amygdala, ILF, IFOF and uncinate fasciculus.

Abnormal voxels for CBS were found in bilateral midbrain regions including brain stem, corticospinal tracts, anterior thalamic radiations, amygdala, thalamus and pallidum. Bilateral abnormal voxels were also found in both lateral ventricles and surrounding regions including cingulate gyrus (anterior and posterior), cingulum cingulate, and the corticospinal tracts. Abnormal voxels were found in the right middle frontal gyrus. Left-hemisphere regions with abnormal voxels included the precentral gyrus, and also regions in the lateral temporal lobe. These included the planum polare, central and parietal operculum cortex, and Heschl’s gyrus. Finally, bilateral abnormal voxels were found in the occipital pole. Significantly abnormal voxels were found in bilateral anterior thalamic radiation, bilateral corticospinal tracts, bilateral pallidum, left occipital pole, bilateral lateral ventricles, right anterior and bilateral posterior cingulate gyrus, bilateral cingulum cingulate, and bilateral supplementary motor cortex.
For PSP, abnormal voxels were found in the entire brainstem and bilateral midbrain regions. Specifically, bilateral corticospinal tract, anterior parahippocampal gyrus, anterior thalamic radiation, thalamus and pallidum all showed bilateral but left-dominant abnormality. The left putamen, but not right, also had abnormal voxels. Additionally, the bilateral (but left greater) precentral gyri also showed abnormal voxels, though less abnormal compared to controls (purple scale) relative to the other regions described so far. Significantly abnormal voxels were found in only in the midbrain; specifically, the brain stem, bilateral anterior thalamic tracts and corticospinal tracts, bilateral thalamus, left putamen, bilateral pallidum, bilateral amygdala, and left posterior parahippocampal gyrus.
Figure 4.3 – Patterns of neural abnormality relative to controls, across aetiologies.

Per subgroup, abnormality maps were averaged (mean) across all patients to generate a mean atrophy map (displayed in the red-purple scale). The scale was set to 15-25% atrophy for all groups. Per subgroup, abnormality maps per subject were entered into a whole brain one-sample t-test analysis to determine which voxels were abnormal relative to zero. The resultant statistical maps were used to generate an outline around the voxels which were significant. The outline is overlayed onto the mean atrophy maps in black. The sagittal slices are displayed at coordinates which show the extent of atrophy specific to each subgroup (note that only post-stroke aphasia is not bilateral). Negative coordinates indicate the left-hemisphere, positive coordinates indicate the right hemisphere.
Voxel-based correlational methodology
Unique neural correlates were found for each of the three behavioural components from the PCA. Figure 4.4 shows the clusters associated with each component, in which tissue abnormality covaries uniquely with factor scores on each component.

Phonology
Performance on the phonology factor was uniquely correlated with voxels across a number of locations in the left hemisphere temporoparietal region (see Figure 4.4A). The cluster overlapped with superior lateral occipital cortex, precuneus cortex, angular gyrus, superior parietal lobule, cingulum cingulate, anterior and posterior supramarginal gyrus, planum temporale, and insula. In terms of white matter, this cluster encompassed parts of the superior and inferior longitudinal fasciculus, the anterior thalamic radiation, and the inferior frontal occipital fasciculus.

Semantics
The semantic component was uniquely correlated with a cluster that encompassed the bilateral anterior temporal lobes (Figure 4.4B). The regions encompassed by the cluster bilaterally included the temporal pole, hippocampus, posterior temporal fusiform cortex, and anterior parahippocampal gyrus. The anterior inferior temporal gyrus was part of the cluster bilaterally, but on the right hemisphere this extended also to the posterior inferior temporal gyrus. Other regions only in the right hemisphere part of the cluster included the anterior temporal fusiform cortex and the inferior frontal occipital fasciculus (the only white matter overlapping with this cluster). On the left hemisphere part of this cluster, the left amygdala and a portion of the frontal orbital cortex were also encompassed by the cluster.

Syntax
The connected speech component uniquely correlated with a cluster overlapping the left lateral ventricle and extending posteriorly and laterally to temporal white matter tracts (Figure 4.4C). The cluster overlapped with anterior and posterior portions of the left lateral ventricle, and with caudate and thalamus. The corticospinal tract, inferior frontal occipital fasciculus, and the inferior longitudinal fasciculus were encompassed by this cluster.
Table 4.3 – Anatomical locations of clusters for phonology, semantics and syntax.

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Location</th>
<th>Extent (voxels)</th>
<th>T</th>
<th>Z</th>
<th>p</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1 (phonology)</strong></td>
<td>Left lateral parietal cortex</td>
<td>4161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lateral occipital cortex sup</td>
<td>5.24 4.83 &lt; .001</td>
<td>-26</td>
<td>-60</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superior parietal lobule</td>
<td>4.89 4.55 &lt; .001</td>
<td>-26</td>
<td>-48</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inferior frontal occipital fasciculus</td>
<td>4.78 4.46 &lt; .001</td>
<td>-30</td>
<td>-36</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td><strong>Factor 2 (semantics)</strong></td>
<td>Bilateral anterior temporal lobes</td>
<td>17870</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left temporal fusiform cortex ant</td>
<td>8.27 6.97 &lt; .001</td>
<td>-30</td>
<td>-8</td>
<td>-36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left temporal pole</td>
<td>6.83 6.02 &lt; .001</td>
<td>-22</td>
<td>16</td>
<td>-32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right temporal fusiform cortex ant</td>
<td>6.55 5.82 &lt; .001</td>
<td>34</td>
<td>-8</td>
<td>-34</td>
<td></td>
</tr>
<tr>
<td><strong>Factor 3 (syntax)</strong></td>
<td>Left hemisphere</td>
<td>2031</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lateral ventricle (posterior peak)</td>
<td>4.57 4.29 &lt; .001</td>
<td>-18</td>
<td>-30</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lateral ventricle (anterior peak)</td>
<td>4.44 4.18 &lt; .001</td>
<td>-6</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inferior frontal occipital fasciculus</td>
<td>4.15 3.93 &lt; .001</td>
<td>-38</td>
<td>-40</td>
<td>-2</td>
<td></td>
</tr>
</tbody>
</table>
Age and scan parameters were not associated with any significant clusters. Lesion size and symptom duration were associated with significant clusters but not explored further since they were added as nuisance covariates. Each t-map is thresholded at $p < .001$, cluster corrected at FWE of $p < .05$; with correction for lesion volume, age, symptom duration and scan parameters. Panels show clusters for (A) phonology, (B) semantic, and (C) syntax principal components. Slices are shown at the coordinate corresponding to the first peak in the cluster. The scales represent the min/max $T$ value for the significance level. Figures are in neurological convention (left is left). Sagittal slices are shown for the left hemisphere.
Discussion

Comparisons of aphasia across aetiologies are scarce, often limited to specific subtypes or linguistic impairments, and rarely compare neural correlates of aphasia in the same neuroimaging analysis format. This study compared language impairments in post-stroke aphasia, primary progressive aphasia, corticobasal degeneration and progressive supranuclear palsy and related the shared underlying aphasic dimensions to unique neural correlates across aetiologies.

Overall, our results showed that variation in language impairments could be captured by three unique principal components reflecting phonological processing, semantics, and syntax. These were associated with abnormal neural tissue in clusters in the left lateral parietal cortex, bilateral anterior temporal lobes, and left hemisphere caudate and nearby white matter. The significance of these findings in the context of other brain-behaviour studies in aphasia is discussed below.

Patterns of linguistic and neural abnormality

This section will summarise the linguistic profiles and neural abnormality maps of each subgroup in the context of the expected deficits and atrophy patterns. First, it is notable that there is considerable variability in abnormality patterns within each subgroup (as evidenced by the red/yellow regions outside the black outline). This could be due to a lack of power as the subgroup sizes range from 11 to 21 cases. However, we believe this shows that grouping cases by their categorical diagnostic label is not necessarily the most appropriate basis for comparison; from Figure 4.2 it can be seen that there is considerable intra- and intergroup heterogeneity on the extracted dimensions. Thus, one would also expect heterogeneity in the patterns of neural abnormality within the proposed diagnostic subgroups. With this in mind, we have discussed whether the expected neural abnormality patterns were broadly captured by the automated identification procedure.

Patients with PSA were impaired on all the subtests of the MLSE relative to controls except semantic association. Within other subtests of the MLSE, the PSA cases were relatively worse for tests of repetition than semantics. This likely reflects the general phenotypic presentations in this cohort which were mostly non-fluent or anomic, rather than having specific semantic impairments (cf., Wernicke’s aphasia). In terms of the subgroup abnormality map, the PSA average abnormality map showed abnormality in the left-hemisphere middle-cerebral artery territory (Phan et al., 2005), replicating the lesion overlap published previously by others with a cohort of PSA that included these 21 cases (Butler et al., 2014, Halai et al., 2017, Schumacher et al., 2019).

In PNFA, performance was significantly impaired on nearly every MLSE subtest compared to controls – the exceptions were word repetition, tests of semantics not requiring verbal response (word-picture matching, semantic association), and irregular word reading. Overall, this suggests that PNFA cases show spared single-word comprehension and object
knowledge, which is in agreement with the characteristic picture for this form of PPA (Gorno-Tempini et al., 2011). PNFA and PSA were the only subgroups to have significantly worse performance than controls for both of the sentence comprehension tests. This implies an impairment in the comprehension of syntactically complex sentences regardless of working memory demands, which is a core feature of PNFA (Hodges and Patterson, 1996, Gorno-Tempini et al., 2004, Grossman et al., 1996). In terms of average abnormality maps, the PNFA subgroup displayed abnormal voxels in a left-dominant pattern encompassing left posterior temporal cortex and underlying white matter. We found the expected pattern of abnormality in left precentral gyrus, middle frontal gyrus, anterior insula, frontal operculum cortex, and caudate, which has been reported for PNFA previously in terms of atrophy (Gorno-Tempini et al., 2004, Ogar et al., 2007) and hypometabolism (Nestor et al., 2003). However, Gorno-Tempini et al. (2004) also found atrophy in the left inferior frontal gyrus and putamen which were not abnormal in the PNFA cases in our study.

Patients with LPA were impaired relative to controls on tests of sentence-level processing, including sentence repetition, writing, spoken picture description, and sentence comprehension I. However, sentence comprehension was not impaired in LPA when the phonological working memory demands were reduced by having picture stimuli to choose from (sentence comprehension II). These results are in alignment with the hypothesis that LPA represents a phonological short-term memory impairment (Gorno-Tempini et al., 2008). LPA also performed significantly worse than controls on naming and non-word repetition. Although post-hoc pairwise comparisons did not survive correction for multiple comparisons, sentence comprehension I and sentence repetition had the lowest average performance in LPA (and no other subgroup performed as poorly on sentence comprehension I). This is in line with the proposed core deficits of LPA, namely impaired single-word retrieval and sentence repetition (Gorno-Tempini et al., 2011). In terms of the neural abnormality, LPA displayed abnormal voxels in an asymmetrical, left dominant, pattern encompassing the superior and posterior temporal lobe, and underlying white matter. In particular, we found the expected pattern of abnormality in the left posterior middle temporal gyrus and posterior superior temporal gyrus, left hippocampus, and left inferior parietal lobe (left angular gyrus), which has been found in terms of atrophy (Gorno-Tempini et al., 2004, Rohrer et al., 2013, Rohrer et al., 2010, Leyton et al., 2012) and hypometabolism (Gorno-Tempini et al., 2008).

Patients with SD were significantly impaired relative to controls, and to patients with PSP, CBS and/or PSA, on tests of semantic knowledge. These tests included naming, spoken-word picture matching, and semantic association, thus demonstrating semantic impairment across tasks with different natures. Only SD cases were impaired on the semantic association subtest, with the average SD score of around 60%, corresponding about 3/4 items correct. This is in alignment with the results of Jefferies and Lambon Ralph (2006) who demonstrated that SD is characterised by consistent semantic impairments across input/output modalities of different tasks, supporting the idea of a loss of amodal (i.e., not
specific to a domain such as verbal output) semantic representations in SD (Lambon Ralph et al., 2017). Compared to the striking semantic impairments in SD, performance on the MLSE subtests for repetition and sentence comprehension with picture stimuli was relatively spared, consistent with previous findings (Rochon et al., 2004, Gorno-Tempini et al., 2004, Peelle et al., 2008). This highlights the selectivity of the semantic impairment in this form of PPA. The average abnormality maps for SD showed the expected pattern of bilateral, but left-dominant, abnormality in anterior temporal lobes and associated underlying white matter (Mion et al., 2010, Lambon Ralph et al., 2017, Acosta-Cabronero et al., 2011, Hodges et al., 1992, Snowden et al., 1989).

In the CBS group, comparisons with healthy controls on MLSE subtests also did not survive correction for multiple comparisons. However, language impairments have been documented for CBS. In particular, people with CBS display impairment in sentence comprehension (Santos-Santos et al., 2016), speech production (Özsancak et al., 2000), single-word and sentence repetition (Burrell et al., 2013), non-word reading (Graham et al., 2003). Writing impairments have also been documented at the single-case level (Gorno-Tempini et al., 2004). To see if these specific impairments were captured by the MLSE, a priori pairwise t-tests were conducted, which showed that CBS patients were significantly impaired relative to controls for sentence comprehension I and writing. Thus, the MLSE captured the expected deficits in these domains, but was not sensitive to other impairments that have been reported for CBS. The average abnormality maps for CBS largely showed the expected pattern of abnormality in left posterior cingulate cortex, bilateral supplementary motor area, left (not bilateral) precentral gyrus, and left occipital pole (Boxer et al., 2006, Lee et al., 2011, Whitwell et al., 2010). However, these studies also found atrophy in CBS extending to prefrontal cortex and premotor areas. The lack of repetition and reading deficits in our findings could relate to not finding abnormality in premotor areas, as these regions have been linked to non-word repetition (Hartwigsen et al., 2013) and reading (Pattamadilok et al., 2016).

In the PSP group, differences in performance across the MLSE did not survive correction for multiple comparisons. However, previous research has highlighted impairments in speech production (Sakai et al., 2002), semantic association (Catricalà et al., 2019), repetition of single words (Burrell et al., 2018), oral reading (Podoll et al., 1991), and writing (Sitek et al., 2015). Therefore, specific a priori pairwise t-tests were conducted, which showed that PSP were impaired relative to controls on reading (words and non-words), writing, and picture description. Thus, the MLSE did not capture impairments in word repetition or semantic association in this subgroup. The average abnormality maps for PSP showed abnormality that was mostly in the brainstem, including the thalamus and pallidum bilaterally, though also including the precentral gyrus. These regions are characteristically atrophied in PSP (Boxer et al., 2006, Cordato et al., 2000, Gröschel et al., 2004), and atrophy in these regions has been shown to have positive predictive value for diagnosis of PSP vs. Parkinson’s disease.
(Stamelou et al., 2011). However, several studies have found frontal atrophy in PSP compared to healthy controls (Brenneis et al., 2004), e.g., frontal opercular cortex (Boxer et al., 2006), which was not evident in our PSP abnormality maps. Furthermore, Barsottini et al. (2007) found that dilation of the third ventricle distinguished PSP from Parkinson’s disease; the third ventricle was not part of the atlas used in our study so we could not confirm whether abnormal voxels were found in the third ventricle of our PSP cases. Overall, we did not find abnormality beyond the midbrain and precentral gyrus for PSP. The MLSE subtests assessing word repetition and semantic knowledge were not impaired in this group, corresponding to the fact that these abilities are possibly supported by regions which were not abnormal relative to controls; there were no abnormal voxels in primary motor cortex, premotor and supplementary motor areas and inferior frontal gyrus, which have been found to be engaged in repetition of words and non-words (Hartwigsen et al., 2013). Since these regions have been shown to be atrophied in PSP in other studies (Brenneis et al., 2004), it's possible that the PSP cases included represent a mild stage of the disease, these MLSE subtests might be insensitive to the specific impairments in PSP, or there might be a lack of statistical power in this group. The semantic association test appears to be insensitive to mild semantic deficits since only the group who performed worse than controls on this test was the SD group who have a pervasive semantic impairment.

Overall, the PSP and CBS groups showed milder deficits on the MLSE, and abnormality restricted to more midbrain and subcortical regions, than the PSA and PPA subgroups. The PSA group showed the most extensive region of abnormal neural tissue, consistent with the stroke aetiology. The PPA subtypes showed the expected pattern of linguistic deficits and pattern of neural abnormality.

**Shared brain-behaviour relationships across aetiologies**

This study is the first to uncover the shared dimensions of variance in linguistic profiles and their neural correlates transdiagnostically in PSA, PPA and movement disorders. Situating the linguistic impairments of these disorders in the same multidimensional space accomplishes two things which have not been possible in previous studies: (1) it allows comparison of relative within-group homogeneity or heterogeneity across groups; (2) it highlights which linguistic profiles are common across groups (i.e., which groups are overlapping in the multidimensional space). Our results show that language impairments across aetiologies can be described in terms of variation in three key dimensions: labelled phonology, semantics and syntax. In general, many of the subgroups showed overlapping impairments on one or more of these dimensions, whilst SD cases occupied a relatively unique part of the multidimensional space signifying their selective and significant semantic impairment.

The first component extracted by PCA captured variation in phonological processing. Tests of repetition at the syllable, word, and sentence level, and also word and non-word reading had strong loadings onto this factor. This is in line with a growing literature in PSA from
previous groups who have identified a similar component (Butler et al., 2014, Halai et al., 2017, Schumacher et al., 2019, Lacey et al., 2017, Mirman et al., 2015a). The cluster associated with worse performance on the phonology component overlapped with the anterior and posterior supramarginal gyrus. This region has an established role in phonological processing, including reading words (Sliwinska et al., 2012, Stoeckel et al., 2009, Price et al., 1997), and auditory processing of non-words (Newman and Twieg, 2001) and syllables (Dehaene-Lambertz et al., 2005). The supramarginal gyrus may have a role as a phonological store (Vigneau et al., 2006) in the phonological loop of short-term memory (Baddeley, 1992). The angular gyrus was also encompassed by this cluster, and this region has been found to have a role in domain-general working memory processes (Humphreys and Lambon Ralph, 2017). This phonology cluster also overlapped with the superior longitudinal fasciculus. This white matter tract is proposed to support the ability to repeat (Berthier et al., 2012), possibly through bidirectional information processing between Wernicke’s area and the precentral cortex (Bernal and Ardila, 2009). The arcuate fasciculus connects the inferior frontal gyrus to the posterior superior temporal gyrus (Parker et al., 2005), which encompasses the planum temporale which was also part of this cluster. The planum temporale has been proposed as a computational hub for processing spectrotetmal patterns (Griffiths and Warren, 2002) including speech, such as syllable perception (Jäncke and Shah, 2002), but also non-language stimuli (Binder et al., 1996). The posterior insula was also encompassed by the phonology cluster. The insula has been found to share connections and activations with diverse language-processing regions (Ardila et al., 2014, Oh et al., 2014). Previous voxel-based morphometry studies in PSA have also related a PCA-derived phonological component to lesions in planum temporale and posterior insula (Halai et al., 2017, Schumacher et al., 2019), but also to lesions in other areas such as middle temporal gyrus and inferior longitudinal fasciculus. Differences between the nature of the phonology components across studies could explain these different neural correlates. Expressive and receptive phonological processing have been shown to have separable neural correlates in more posterior and anterior locations, respectively (Alyahya et al., in press). In our study, the phonology factor included strong loadings from tests of oral reading and sentence repetition, which have strong components of expressive phonology. The phonology components in Halai et al. (2017), Butler et al. (2014), and Schumacher et al. (2019), included more tests capturing receptive phonology (e.g., phonological discrimination of minimal pairs), which possibly resulted in the more anterior neural correlates of phonology in these studies.

The second component extracted by PCA in our study captured variance in semantics and was related to a large asymmetric cluster located in bilateral anterior temporal lobes (ATL). This is in line with previous studies in PSA which have found the left ATL to be associated with a principal dimension capturing semantic impairments (Butler et al., 2014, Halai et al., 2017, Schumacher et al., 2019, Mirman et al., 2015a). In our study, the semantic cluster encompassed bilateral temporal fusiform cortex, parahippocampal gyri, and inferior temporal
gyrus. Left and right anterior fusiform function is associated with expressive verbal semantic tasks (spoken picture naming) and non-verbal semantic association tasks, respectively (Mion et al., 2010). The posterior inferior temporal gyrus has been shown to be involved in reading words vs. non-words (Fiez et al., 1999; Hagoort et al., 1999), and verbal semantic knowledge retrieval in semantic association (Booth et al., 2002). The right IFOF was also encompassed by the semantic cluster, and this white matter tract has been found to have a role in semantic processing (Han et al., 2013). Overall, these results converge with previous findings from transcranial magnetic stimulation (Pobric et al., 2007), intraoperative stimulation studies (Duffau et al., 2008), imaging studies of SD (Mion et al., 2010), functional MRI (Vigneau et al., 2006, Visser et al., 2010, Binder et al., 2009), and computational models (Ueno et al., 2011).

The third component extracted by PCA had strong loadings from tests capturing both comprehension and production of sentences. The tests loading onto this component generally involved connected speech (listening to and/or producing it, verbally or non-verbally (i.e., in writing)). Since language impairments at the level of connected speech capture variance in syntactic processing (e.g., of complex grammatical structure), we labelled this component syntax. However, there were moderate loadings from tests of non-word reading and non-word repetition, which suggest that this component should be replicated and explored further in future studies. The syntax cluster overlapped the left lateral ventricle, left caudate, and nearby temporal lobe white matter tracts. These regions have been associated with functions including speech generation, reading and writing. Halai et al. (2017) found a principal component reflecting speech quanta, which was associated with unique neural correlates in left hemisphere caudate nucleus, in addition to pre-central gyrus, insula, and putamen. The caudate has been found to have a role in word generation (in response to a rhyming word or a semantic category) (Crosson et al., 2003). Others have found that the left caudate is associated with switching between languages in bilinguals (Crinion et al., 2006). The cluster encompassed the IFOF and ILF, which are part of the ventral language route (Parker et al., 2005, Bajada et al., 2015) and damage to the IFOF has been associated with reading and writing impairments (Tomasino et al., 2015). In general, the PSP and CBS were relatively spared on the phonology and semantics dimensions but displayed significant impairments on sentence-level tests (e.g., sentence comprehension and writing). As such, we could speculate that this syntax component-cluster may be driven by the movement disorders, but this needs further exploration in future studies.

In summary, our results and those outlined above show that the phonology and semantics components are highly reproducible dimensions of variance in aphasia. Thus, these dimensions might reflect core “primary systems” (Patterson and Lambon Ralph, 1999, Ueno et al., 2014, Woolllams et al., 2018) of language which are affected to different degrees in different patient populations.
Conclusions and future directions

A key advantage of this study was analysing neural abnormality in the same way across stroke and neurodegenerative aetiologies. Overall, the patterns of neural abnormality per subgroup matched the expected patterns of atrophy and hypometabolism reported in the literature. This suggests that the automated lesion identification procedure (Seghier et al., 2008) was capable of detecting abnormality due to neurodegenerative atrophy instead of stroke lesions. Future research could explore the generalisability of methods traditionally employed to study dementia to stroke, for example cortical thickness analysis (Jin Thong et al., 2014).

We used VBCM to relate continuous voxel intensity values to the continuous principal dimensions extracted by PCA. This method avoids having to binarise voxel integrity, as in VLSM. However, a potential limitation of univariate methods like VBCM is that they do not account for neural correlates which are related to more than one cognitive function. Many brain regions are associated with multiple functions, such as the anterior superior temporal gyrus and posterior inferior temporal gyrus which have each been associated with both phonological and semantic processing (Vigneau et al., 2006). Multivariate decoding methods (Halai et al., in press, Murley et al., 2019) are required to uncover these more inter-related brain-behaviour relationships.

To conclude, the present study is the first to compare directly linguistic and neural abnormalities across neurodegenerative and stroke aetiologies. Crucially, we compared language impairments on the same transdiagnostic assessment tool, situated these impairments in the same, shared multidimensional space, and analysed neural abnormality across aetiologies in the same voxel-based morphometry analysis. By unifying these methodological approaches were we able to extract brain-behaviour relationships common to aphasia across aetiologies.
Chapter 5 - PCA-squared: mapping graded differences in posterior cortical atrophy and typical Alzheimer’s disease using Principal Component Analysis

Ruth U. Ingram¹, Keir X.X. Yong², Ajay D. Halai³, Gorana Pobric¹,
Sebastian J. Crutch² & Matthew A. Lambon Ralph³

¹Division of Neuroscience and Experimental Psychology,
School of Biological Sciences, University of Manchester, UK
²Dementia Research Centre, UCL Institute of Neurology, London, UK
³Department of Clinical Neurosciences, University of Cambridge & MRC Cognition & Brain Sciences Unit, Cambridge, UK

Statement of contribution

Ruth Ingram analysed all the data presented in this chapter. Keir Yong and Sebastian Crutch collected the behavioural data. Matt Lambon Ralph, Ajay Halai, and Gorana Pobric provided guidance and support throughout the analysis, and feedback on drafts of the write up.
Abstract

Alzheimer’s disease (AD) pathology is known to cause a heterogeneous collection of clinical phenotypes, which are recognised in current diagnostic criteria as typical and atypical subtypes, or variants. One atypical variant, known as posterior cortical atrophy (PCA), is characterised by impairment in visual processing, in contrast to episodic memory impairment in typical AD. PCA has been proposed to comprise two subtypes reflecting differential impairment in visuospatial and visuoperceptual processing, respectively. However, the nature of the relationships between typical and atypical AD, and within the heterogeneity of PCA, might be better conceptualised as one or multiple phenotypic spectra, instead of subtypes (Warren et al., 2012, Stopford et al., 2008, Ridgway et al., 2012, Lambon Ralph et al., 2003, Migliaccio et al., 2009, Fitzpatrick et al., 2019). We applied principal component analysis to neuropsychological data from a large cohort of typical AD and PCA (total N = 155) on an extensive cognitive and visual test battery. The principal component analysis for PCA extracted three components which were interpreted as general cognitive status, visuoperceptual processing and visuospatial processing. When PCA cases’ factor scores were plotted on the visuoperceptual and visuospatial components, there was no evidence of distinct groups of cases with selective impairments in these visual domains. Instead, we found graded variation along both dimensions. The principal component analysis for AD extracted two components, reflecting cognitive status and general visual processing. Plotting AD cases’ factor scores on these two components showed that some displayed relatively impaired visual processing compared to general cognitive status. When the PCA cases’ factor scores on the general cognitive status and general visual processing components were plotted alongside the AD cases, these AD cases with visual impairments overlapped completely with PCA cases who showed milder visual impairments. Thus, we found evidence of PCA-like visual impairments in this cohort of typical AD. Both PCA and typical AD cases varied gradedly along the visual processing dimension, implying a spectrum of visual processing impairments between the typical AD, PCA-like AD, mild PCA and more severe PCA. Our results therefore support the conceptualisation of (a) two phenotypic spectra in PCA rather than discrete visuoperceptual/visuospatial subtypes, and (b) a spectrum of visual impairments in typical and atypical AD.
Introduction

Alzheimer’s disease (AD) pathology is known to cause a heterogeneous collection of clinical phenotypes. The current diagnostic criteria delineate typical and atypical presentations, which are conceptualised as distinct variants of AD (Dubois et al., 2014). However, many groups have proposed that the relationship between these variants might be better described as a phenotypic continuum (Warren et al., 2012, Stopford et al., 2008, Ridgway et al., 2012, Lambon Ralph et al., 2003, Migliaccio et al., 2009, Fitzpatrick et al., 2019), instead of discrete subtypes. The aim of this study was to evaluate whether this relationship comprises categorical differences (implying discrete subtypes) or graded variation (implying one or multiple phenotypic spectra).

Typical AD is defined by impairment in hippocampal-dependent episodic memory (Varma et al., 1999). Atypical AD is characterised by a principal impairment in a different cognitive domain, with relative sparing of episodic memory (Dubois et al., 2014). The three most common variants of atypical AD are termed frontal, logopenic, and parietal (Dubois et al., 2007). Some studies have reported categorical differences between variants of AD (Mendez et al., 2002, Tsai et al., 2011, Tang-Wai and Mapstone, 2006), whereas other studies have found considerable overlap in key features such as the distribution of grey matter atrophy (Migliaccio et al., 2009, Lehmann et al., 2011b, Ridgway et al., 2012, Ossenkoppele et al., 2015, Lehmann et al., 2010), and the constellation of impairments in shared cognitive domains (Crutch et al., 2012b, Migliaccio et al., 2009). Still other studies have reported cases of typical AD that deviate mildly (i.e., not enough to no longer meet the criteria for typical AD) along non-memory cognitive domains, such as semantics/verbal processing, and visuospatial processing (Stopford et al., 2008, Snowden et al., 2007, Lambon Ralph et al., 2003, Caine and Hodges, 2001, Kanne et al., 1998, Price et al., 1993). The overlap between presentations, and atypical-like presentations within typical AD, suggest that the relationship between typical and atypical AD might be better conceptualised as comprising one or more phenotypic spectra.

This same conceptual debate is also relevant within the parietal variant of AD, also known as posterior cortical atrophy (PCA, (Benson et al., 1988)), for which two subtypes have been proposed (Ross et al., 1996). This was motivated by the separation of healthy visual processing into separate streams (Crutch et al., 2012a). Ungerleider and Mishkin (1982) described the dissociation of higher-order visual processing into dorsal “where” and ventral “what” streams in primates. Goodale and Milner (1992) proposed a model whereby the dorsal “where” stream processes visuospacial information, whilst the ventral “what” stream processes visuoperceptual information. Studies investigating these proposed subtypes of PCA have found that cases showing relatively selective ventral/dorsal symptoms have been associated with neural changes in occipito-temporal and occipito-parietal cortices respectively (Migliaccio et al., 2012, Tsai et al., 2011, Tang-Wai et al., 2004).
However, as with comparisons of typical and atypical AD, some studies have reported a degree of overlap between proposed dorsal and ventral subtypes of PCA (Crutch et al., 2017). Tang-Wai et al. (2004) and Tsai et al. (2011) both note that although they found clusters/groups of dorsal/ventral presentations, there was a degree of overlap. Others have found such substantial overlap that it precluded a conclusion supporting subtypes of PCA at all; Lehmann et al. (2011a) divided their PCA cases into Space (dorsal) and Object (ventral) groups based on composite scores from an extensive test battery of visual processing, yet they found considerable overlap in the distribution of reductions in cortical thickness in the two groups. Thus, Lehmann et al. (2011a) and others (Crutch et al., 2017) proposed a re-conceptualisation of the heterogeneous presentations of PCA as varying along one or multiple phenotypic spectra, instead of reflecting discrete subtypes.

In summary, for the varied phenotypes associated with AD, and for the potential dichotomy in PCA, there is a conceptual question: are there subtypes or spectra? To answer this question, it is important to adopt a methodology capable of capturing continuous variation. Principal component analysis is a statistical procedure which extracts a number of factors to best explain the variation in a set of data, e.g., neuropsychological data from different patient groups. The extracted continuous components (or factors) can be thought of as the principal behavioural dimensions which describe the phenotypic space of a disorder. Principal component analysis generates factor scores for every patient (see Methods) which can be treated as coordinates within the phenotypic multidimensional space (Armstrong et al., 2000, Butler et al., 2014, Martin et al., 1986). For example, Armstrong et al. (2000) applied principal component analysis to neuropathological data from cases of sporadic and familial AD (both early- and late-onset) to test the hypothesis that these are separable subtypes of AD. Plotting individual cases onto pairs of their principal factors showed extensive overlap of all groups and continuous variation along all factors, thus providing evidence against the Subtype hypothesis (Ritchie and Touchon, 1992).

The aim of our study was therefore to apply this alternative principal component analysis approach to neuropsychological data from cases of typical AD and PCA to explore whether there was evidence of continuous variation 1) between typical AD and PCA as an example of atypical AD, and 2) within PCA itself. We applied principal component analysis separately to our cohorts of typical AD and PCA to establish the unique multidimensional phenotypic space of each variant of AD. We then visualised the variation amongst individual cases in each group by plotting them within the multidimensional space defined by the extracted factors for that group. We additionally plotted cases of PCA within the multidimensional space of typical AD to assess whether there were graded differences between these groups.
Materials and Methods

Patients

Ninety-four people with posterior cortical atrophy were recruited on the basis of showing, at some time during their clinical history, posterior cognitive impairment with preservation of episodic memory. Specifically, all people with posterior cortical atrophy met the Dementia Research Centre behavioural criteria of relatively preserved episodic memory (>5th percentile on a Recognition Memory Test (Warrington, 1984, Warrington, 1996)), along with neuropsychological deficits (<5th percentile) in at least two of the following four posterior functions: (1) object perception (Object Decision test) and (2) space perception (Number Location test) from the Visual Object and Space Perception battery (Warrington and James, 1991); (3) calculation (Graded Difficulty Arithmetic test (Jackson and Warrington, 1986)); and (4) spelling (Graded Difficulty Spelling test (Baxter and Warrington, 1994)). Furthermore, all people with posterior cortical atrophy met the criteria proposed by Tang-Wai et al. (2004), and Mendez et al. (2002). At the time of testing some people with posterior cortical atrophy had more generalised cognitive impairments including mild episodic memory impairments. Exclusion criteria were evidence of non-Alzheimer’s disease dementia (e.g., dementia with Lewy bodies, or cortico-basal degeneration) in clinical assessment and investigation.

Sixty-one people with typical Alzheimer’s disease were recruited on the basis of meeting the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable AD with recently proposed revisions (Dubois et al., 2007, McKhann et al., 1984). Furthermore, these cases additionally demonstrated significant episodic memory impairment (<5th percentile on verbal and visual Recognition Memory Tests (Warrington, 1984, Warrington, 1996)).

The posterior cortical atrophy and typical Alzheimer’s disease cases have been published elsewhere (Firth et al., 2019, Lehmann et al., 2011a, Lehmann et al., 2011b, Ridgway et al., 2012, Ryan et al., 2014, Yong et al., 2014b, Yong et al., 2014a, Primativo et al., 2017). All patients provided informed consent under approval from NRES Committee London, Queen Square. Independent t-tests were conducted to assess significant differences in Mini-mental State Examination (MMSE) (Folstein et al., 1975) scores and demographic variables between groups.

Neuropsychological assessments

The tests included are shown in Table 5.2. Briefly, we measured attention and executive function (Graded Difficulty Arithmetic (Jackson and Warrington, 1986), Cognitive Estimates (Shallice and Evans, 1978), Digit span (forwards and backwards (Wechsler, 1981)), letter “A” Cancellation (Willison and Warrington, 1992), language ability (Concrete Synonyms (Warrington et al., 1998), graded difficulty naming from verbal description, spelling (Graded-difficulty Spelling Test (Baxter and Warrington, 1994)), and short-term recognition memory
A detailed battery of visual tests was also administered: early visual perception (hue discrimination (CORVIST (James et al., 2001)), shape discrimination (Efron, 1968), figure/ground separation (VOSP – Visual Object and Space Perception battery (Warrington and James, 1991)), and crowding (CORVIST (James et al., 2001))); visuoperceptual abilities (usual/unusual views (Warrington and James, 1988), object decision and fragmented letters (VOSP)); visuospatial abilities (dot counting and number location (VOSP)).

**Behavioural analysis**

All raw behavioural scores were converted to percentages. For measures without a fixed maximum score, scores were converted to a percentage of the maximum score within each cohort. The adequacy of the sample size for each principal component analysis was determined using the Kaiser-Meyer-Olkin measure of sampling adequacy.

**Imputation and component selection**

In the interest of retaining as much information (patients and tests) as possible, missing data were imputed using probabilistic principal component analysis (PPCA) (Ilin and Raiko, 2010). In this procedure, the number of components for PPCA to extract must be pre-specified. A k-fold cross-validation approach was used to choose the component-solution (i.e., number of components) with the lowest root mean squared error (RMSE) for held-out cases over 1000 permutations (Ballabio, 2015). This approach was also used to select the optimal number of components for subsequent principal component analysis using the imputed (i.e., complete) dataset.

**Principal component analysis**

We applied principal component analysis separately to the data from the typical AD and PCA cohorts to establish the separate multidimensional space of each presentation, without any potential influence from the other cohort. The principal components (or factors) extracted by principal component analysis reflect orthogonal linear combinations of the tests entered into the principal component analysis. The factor loadings which are produced for each test per factor represent the weighting of that test onto that factor. Each patient is given a factor score per extracted factor. For each factor, a patient’s factor score is the sum of their original test scores multiplied by the factor loadings of those tests onto that factor. The factor scores are normalised such that zero represents the global average for all patients (i.e., they not normalised to control scores).

Principal component analysis overcomes the collinearity inherent in test scores by decomposing variation such that relatively pure aspects of each test (e.g., visuospatial processing) can be amalgamated in the form of a principal component. The cognitive interpretability of these components can be aided by applying varimax rotation, which results in each test loading maximally onto as few components as possible. Having interpreted the
extracted components based on the test loadings, we qualitatively compared what this implied about the underlying structure of variation in PCA and typical AD.

**Correlations with MMSE as a proxy for disease severity**

All extracted components were checked for significant correlations with MMSE scores as a proxy control for disease severity. To do this, factor scores from the resultant analyses were correlated with the MMSE scores, either per diagnostic group or as an entire group. Significant correlations with MMSE were taken to indicate that the component was being driven by general cognitive status/disease severity.

**Projecting PCA into the AD multidimensional space**

After establishing the unique multidimensional space for typical AD, we explored whether there were any regions of this space which would also be occupied by PCA cases. Factor scores are derived by taking the sum of the normalised raw behavioural scores, which had been transformed according to the coefficients of each behavioural test on the extracted components. Therefore, the PCA normalised behavioural scores could be transformed according to the AD-coefficients and summed to give PCA pseudo-factor scores. This achieved two purposes: firstly, it allowed us to account for the fact that the test battery is tailored towards PCA instead of typical AD (a large proportion of the test battery is dedicated to detailed visual assessments, whilst there are relatively few tests of episodic memory). Secondly, it allowed us to visualise potential overlap of typical AD and PCA within the extracted ‘space’ of AD in order to answer our question of whether there is graded variation between typical AD and atypical AD (PCA) suggestive of phenotypic spectra, or if there is evidence of distinct boundaries between typical AD and PCA in the AD ‘space’ suggestive of categorical subtypes.

**Results**

Demographic details are summarised in Table 5.1. There were no significant differences between AD and PCA groups in either age at onset ($t_{(110)} = -1.885, p = .062$) or symptom duration ($t_{(106)} = 1.502, p = .136$). MMSE scores were not significantly different between typical AD and PCA ($t_{(145)} = -0.826, p = .070$) (note that this t-test was conducted on the raw data, i.e., before imputation, and therefore only 56 AD and 91 PCA had a score on the MMSE).

**Demographics**

**Table 5.1 – Demographic details per diagnostic group.**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Total N (F)</th>
<th>Age at onset</th>
<th>Symptom duration (years)</th>
<th>MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>61 (23)</td>
<td>57.2 (6.3)</td>
<td>6.2 (2.9)</td>
<td>19.7 (4.8)</td>
</tr>
<tr>
<td>PCA</td>
<td>94 (59)</td>
<td>60.0 (8.1)</td>
<td>5.3 (2.5)</td>
<td>21.3 (5.1)</td>
</tr>
</tbody>
</table>

Age at onset, symptom duration and MMSE score are presented as mean (SD).
Posterior cortical atrophy

The adequacy results revealed that the posterior cortical atrophy sample size was adequate for principal component analysis (Kaiser-Meyer-Olkin = 0.868). This produced a 3-factor rotated solution which accounted for 61.1% of variance in patients’ performance (F1 = 23.2%, F2 = 21.3%, F3 = 16.6%). The factor loadings of each behavioural assessment onto the extracted components are shown in Table 5.2.

Measures that loaded heavily onto the first factor were tests of visuoperceptual function (usual/unusual views, object decision and fragmented letters), early visual perception (crowding, hue discrimination, shape discrimination and figure/ground discrimination), and short-term recognition memory for words. A common feature of these tests is identification of stimuli based on their perceptual features; hence we called this factor ‘Visuoperceptual processing’.

Measures that loaded heavily onto the second factor were tests of recognition of faces, language abilities (synonym judgement, naming, spelling), and attention/executive tasks (digit span forwards and backwards, arithmetic, and cognitive estimates). This component had a moderate and significant correlation with MMSE (r = 0.602, p < .001). This correlation, coupled with the diverse nature of the tests which loaded onto this factor, led us to call this factor ‘General cognitive status’.

Measures that loaded heavily onto the third factor were tests of visuospatial processing (dot counting and number location) and letter cancellation. Moderate loadings were also found for tests of early visual processing and identifying fragmented letters. This component had a weak but significant correlation with MMSE (r = 0.387, p < .001). These tests all require scanning of the visual field/visual search, and abstraction of visuospatial relationships between stimuli, hence we called this factor ‘Visuospatial processing’.
<table>
<thead>
<tr>
<th>Test</th>
<th>Visuoperceptual processing</th>
<th>General cognitive status</th>
<th>Visuospatial processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usual views</td>
<td>0.891</td>
<td>0.055</td>
<td>0.263</td>
</tr>
<tr>
<td>Unusual views</td>
<td>0.871</td>
<td>-0.039</td>
<td>-0.012</td>
</tr>
<tr>
<td>VOSP object decision</td>
<td>0.860</td>
<td>0.012</td>
<td>0.142</td>
</tr>
<tr>
<td>VOSP fragmented letters</td>
<td>0.650</td>
<td>0.117</td>
<td>0.463</td>
</tr>
<tr>
<td>CORVIST hue discrimination</td>
<td>0.611</td>
<td>0.216</td>
<td>0.227</td>
</tr>
<tr>
<td>Crowding (time)</td>
<td>0.657</td>
<td>0.217</td>
<td>0.436</td>
</tr>
<tr>
<td>VOSP figure/ground</td>
<td>0.538</td>
<td>0.009</td>
<td>0.469</td>
</tr>
<tr>
<td>Efron shape discrimination</td>
<td>0.460</td>
<td>0.125</td>
<td>0.415</td>
</tr>
<tr>
<td>Recognition (words)</td>
<td>0.775</td>
<td>0.116</td>
<td>0.222</td>
</tr>
<tr>
<td>Recognition (faces)</td>
<td>-0.110</td>
<td></td>
<td>0.557</td>
</tr>
<tr>
<td>Graded Difficulty Naming</td>
<td>0.184</td>
<td></td>
<td>0.786</td>
</tr>
<tr>
<td>Concrete synonyms</td>
<td>0.140</td>
<td></td>
<td>0.769</td>
</tr>
<tr>
<td>Baxter spelling</td>
<td>0.131</td>
<td></td>
<td>0.781</td>
</tr>
<tr>
<td>Graded Difficulty Arithmetic</td>
<td>-0.017</td>
<td></td>
<td>0.749</td>
</tr>
<tr>
<td>Digit span (forwards)</td>
<td>0.143</td>
<td></td>
<td>0.690</td>
</tr>
<tr>
<td>Digit span (backwards)</td>
<td>-0.068</td>
<td></td>
<td>0.805</td>
</tr>
<tr>
<td>Cognitive estimates</td>
<td>-0.196</td>
<td></td>
<td>-0.691</td>
</tr>
<tr>
<td>Cancellation (N correct)</td>
<td>0.384</td>
<td></td>
<td>0.595</td>
</tr>
<tr>
<td>Cancellation (time)</td>
<td>0.249</td>
<td></td>
<td>0.540</td>
</tr>
<tr>
<td>VOSP number location</td>
<td>0.242</td>
<td></td>
<td>0.148</td>
</tr>
<tr>
<td>VOSP dot count (N correct)</td>
<td>0.192</td>
<td></td>
<td>0.317</td>
</tr>
<tr>
<td>VOSP dot count (time)</td>
<td>0.176</td>
<td></td>
<td>0.033</td>
</tr>
</tbody>
</table>

Factor loadings per test on the three extracted factors. Factor loadings larger than 0.4 are shaded in grey and loadings larger than 0.5 are additionally in bold.
We then treated these extracted factors as core dimensions along which the PCA cases varied. To visualise variation along these dimensions, all PCA cases were plotted onto pairs of extracted dimensions (Figure 5.1). In panels A-C of this figure, each quadrant of every graph is populated with cases and there are no obvious groups that seem to ‘sit’ apart from other cases. In fact, these plots show continuous variation along all three extracted dimensions, including the visuospatial and visuoperceptual dimensions (panel Figure 5.1C).

To explore these dimensions further, we selected key individuals with opposing factor scores on Factor 1 (visuoperceptual processing) and Factor 3 (visuospatial processing) and plotted their test scores on tests with strong loadings onto these respective factors (Figure 5.1D-E). From panel Figure 5.1E it can be seen that high/low factor scores on Factor 1 (visuoperceptual processing) are associated with good/poor performance on Usual Views, respectively. High/low scores on Factor 3 (visuospatial processing) are associated with good/poor performance on VOSP dot counting. These are examples of individuals selected to illustrate a particularly extreme sub-region of the principal component ‘space’, and therefore they appear to show a categorically dissociable pattern of performance.

However, in panel C, both the top left and bottom right quadrants are populated with cases which vary continuously from each other along both dimensions. If one only considered cases such as those in panel D (e.g., by applying selection criteria), this might point towards dissociable ventral/dorsal presentations of PCA. However, the scattering of cases in these quadrants shows that there is actually continuous variation in both visuoperceptual and visuospatial processing, with no discrete ‘groups’ of cases evident in this cohort.
Figure 5.1 – Graded variation along principal dimensions for posterior cortical atrophy.

Factor scores for all cases on pairs of extracted dimensions. The origin of each plot is the mean of all cases. Panel D: four cases (coloured diamonds) were selected to illustrate how sampling from the extreme ‘corners’ of the multidimensional space can give a false impression of category-like differences between presentations of posterior cortical atrophy. The cases were labelled (moving clockwise from the top) F1_high, F3_high, F1_low, and F3_low. Panel E shows their test scores (expressed as a % of maximum score) on tests representing Factor 1 (Usual Views) and Factor 3 (VOSP dot counting – number correct), to illustrate how cut-off criteria could isolate these cases. Vertical grey bars show normative performance of healthy controls (mean +/- SD).
Alzheimer's disease

The adequacy results revealed that the Alzheimer's disease sample size was adequate for principal component analysis (Kaiser-Meyer-Olkin = 0.800). This produced a 2-factor rotated solution which accounted for 44.0% of variance in patients’ performance (F1 = 23.8%, F2 = 20.2%). The factor loadings of each behavioural assessment onto the extracted components are shown in Table 5.3.

Measures that loaded heavily onto the first factor were tests of short-term recognition memory (words and faces), visuoperceptual ability (usual and unusual views), working memory (arithmetic, digit span backwards, cognitive estimates), language (naming, concrete synonyms, spelling), hue discrimination and visuospatial processing (number location). This factor was weakly but significantly correlated with MMSE score (r = 0.388 p < .002). This correlation, coupled with the diverse nature of the tests which loaded onto this factor, led us to call this factor ‘General cognitive status’.

Measures that loaded heavily onto the second factor were direct tests of visuospatial ability (number location, dot counting), visuoperceptual processing (fragmented letters), early visual processing (figure/ground separation, shape discrimination, crowding), and attention/executive function (digit span, letter cancellation). We called this factor ‘Visual processing’ since it seemed to capture most of the visual tests. It could be the case that visual processing impairments in our cohort could be related to more severe typical AD. However, the visual processing factor was not significantly correlated with MMSE (r = -0.129, p = 0.321).

We then treated these extracted factors as core dimensions along which the AD cases varied. To visualise variation along these dimensions, factor scores for all AD cases were plotted along the two extracted dimensions (Figure 5.2A). As was the case for PCA, Figure 5.2A shows that there were no obvious groups of AD cases which ‘sit’ apart from other cases in any quadrant. This suggests that within this cohort of typical AD there is important graded variation in cognitive status and visual processing, but no evidence of categorical subtypes within typical AD.
Table 5.3 – Principal component analysis for typical Alzheimer’s disease.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Test</th>
<th>General cognitive status</th>
<th>Visual processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visuoperceptual</td>
<td>Usual views</td>
<td>0.717</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>Unusual views</td>
<td>0.733</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td>VOSP object decision</td>
<td>0.089</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>VOSP fragmented letters</td>
<td>0.389</td>
<td>0.708</td>
</tr>
<tr>
<td>Early visual</td>
<td>CORVIST hue discrimination</td>
<td>0.488</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>Crowding (time)</td>
<td>0.012</td>
<td>0.438</td>
</tr>
<tr>
<td></td>
<td>VOSP figure/ground</td>
<td>0.160</td>
<td>0.464</td>
</tr>
<tr>
<td></td>
<td>Efron shape discrimination</td>
<td>0.057</td>
<td>0.462</td>
</tr>
<tr>
<td>Short-term memory</td>
<td>Recognition (words)</td>
<td>0.740</td>
<td>-0.126</td>
</tr>
<tr>
<td></td>
<td>Recognition (faces)</td>
<td>0.430</td>
<td>0.217</td>
</tr>
<tr>
<td>Language</td>
<td>Graded Difficulty Naming</td>
<td>0.873</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>Concrete synonyms</td>
<td>0.705</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>Baxter spelling</td>
<td>0.556</td>
<td>0.244</td>
</tr>
<tr>
<td>Attention/Executive function</td>
<td>Graded Difficulty Arithmetic</td>
<td>0.671</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td>Digit span (forwards)</td>
<td>0.079</td>
<td>0.404</td>
</tr>
<tr>
<td></td>
<td>Digit span (backwards)</td>
<td>0.421</td>
<td>0.592</td>
</tr>
<tr>
<td></td>
<td>Cognitive estimates</td>
<td>-0.671</td>
<td>-0.299</td>
</tr>
<tr>
<td></td>
<td>Cancellation (N correct)</td>
<td>0.173</td>
<td>0.433</td>
</tr>
<tr>
<td></td>
<td>Cancellation (time)</td>
<td>0.173</td>
<td>0.770</td>
</tr>
<tr>
<td>Visuospatial</td>
<td>VOSP number location</td>
<td>0.515</td>
<td>0.614</td>
</tr>
<tr>
<td></td>
<td>VOSP dot count (N correct)</td>
<td>0.223</td>
<td>0.701</td>
</tr>
<tr>
<td></td>
<td>VOSP dot count (time)</td>
<td>0.001</td>
<td>0.754</td>
</tr>
</tbody>
</table>

Factor loadings per test on the two extracted factors. Factor loadings greater than 0.4 are shaded in grey, with loadings greater than 0.5 additionally in bold.
If worse visual processing was a consequence of later/more severe typical AD, then this would preclude the bottom right quadrant capturing any cases (i.e., all cases in the bottom half of the graph would be associated with worse cognitive status). However, as there are cases of typical AD in the bottom right quadrant of Figure 5.2A, this could represent a phenotype which is similar to PCA, i.e., relatively impaired visual processing in the context of relatively spared performance in other cognitive domains. Importantly, these AD cases do not occupy a discrete region of the AD ‘space’ (rather, they are scattered across the bottom right quadrant), implying that they are gradedly different from other cases of typical AD.

The cases in this bottom right quadrant appear to be atypical-like presentations of typical AD, i.e., PCA-like AD. This finding is in agreement with those who have found similar atypical-like presentations within typical AD, but it remains to be established whether there is a categorical boundary between PCA-like typical AD and ‘true’ PCA, or if they represent different regions along a spectrum of visual impairment in AD.

**Projecting PCA into AD-space**

To explore this PCA-like AD presentation and its relation to PCA, we placed PCA within the two-dimensional ‘space’ of typical AD; this enabled the comparison of which regions of the space (cf. behavioural presentations) were represented in each cohort, and if these regions were occupied uniquely (suggestive of categorical differences) or if there was overlap (suggestive of a phenotypic spectrum). This was achieved by plotting PCA pseudo-factor scores alongside typical AD factor scores on the extracted AD dimensions (Figure 5.2B).

The key observation from plotting AD and PCA in the same ‘space’ was the degree of overlap of cases on both dimensions. The full range of cognitive status is sampled in both patient groups. Although the extremes of the visual processing dimension are occupied by typical AD (top) and PCA (bottom) uniquely, the rest of this dimension is populated by cases from both groups, namely the PCA-like AD and PCA cases. In other words, visual processing ability is highly overlapping in PCA and atypical-like typical AD.

This was explored further by examining the test scores of illustrative cases in key positions along the visual processing dimension (Figure 5.2C-D). Illustrative cases were selected to have matched factor scores for general cognitive status, to aid isolation and interpretation of variation on the visual processing factor. Figure 5.2D shows that the AD case with the lowest Visual processing factor score out of the AD cohort (AD_low) has equivalent performance on VOSP number location with an ‘average’ PCA case (PCA_mid), whilst both perform better than the PCA case with the lowest Visual processing factor score of the PCA cohort (PCA_low).
Figure 5.2 – Graded variation along a spectrum of visual processing impairments in typical AD and posterior cortical atrophy.

Factor scores for cases are plotted on the two extracted principal dimensions. Panel A shows the heterogeneity in cognitive status and in visual processing in typical AD. Panel B: AD factor scores and PCA pseudo-factor scores plotted on the AD dimensions. Panels C and D: AD and PCA illustrative cases’ factor scores and test scores (mean +/- SD) demonstrating a spectrum of visual impairment between typical AD and PCA. Vertical grey bars show normative performance of healthy controls. In Panels C and D, the AD_high (light green) case displays relatively spared visual processing, whilst PCA_low (dark blue) displays severely impaired visual processing. In Panels C and D, AD_low and PCA_high (dark green and light blue, respectively) display overlapping impairments in visual processing. Thus, these four cases represent points on a spectrum of visual impairment across typical AD and PCA, which is shown in full in Panel B.
Discussion

In this study we sought to address a conceptual question of theoretical and clinical relevance to typical/atypical Alzheimer’s disease, and to the potential phenotypic dichotomy in PCA: are there subtypes or spectra? Using principal component analysis to explore continuous variation in typical AD and PCA, we have shown that there are graded differences in visual processing between typical AD and PCA, and furthermore that there are graded differences in visuoperceptual and visuospatial processing within PCA itself. Both results highlight the potential utility of conceptualising variation in these forms of Alzheimer’s disease as multiple phenotypic spectra instead of categorically dissociable subtypes.

Our results suggest that behavioural variation in AD and in PCA cannot be explained by a single general disease severity factor, since at least two principal components were extracted for each group. If behavioural variation in these forms of Alzheimer’s disease could simply be explained by general disease severity then the principal component analysis would have extracted a single factor. This result is in contrast to others who have found a single severity-governed factor which explained a large proportion of variation in AD (Lambon Ralph et al., 2003, Salthouse and Becker, 1998). This could be because the groups in our study were relatively homogeneous in terms of disease severity within each group (even though some PCA cases had mild episodic memory impairments at the time of testing), in contrast to Lambon Ralph et al. (2003) who were investigating the longitudinal profile of cases with diverse disease stages. Salthouse and Becker (1998) applied common factor analysis to a neuropsychological test battery on AD, and found that a single severity factor accounted for 75% of variance in their data. Common factor analysis analyses only the variance which is common to all variables in order to find the fewest number of factors to explain all common variance, whereas PCA analyses all variance in a set of variables (both common and unique) to find the number of factors that best explains all of this variance (Kim, 2008). As a result, the disparity between our multi-factor solution and their single-factor solution could be attributable to the nature of the factor analysis employed. Overall, our results suggest that in typical AD and in PCA there is important orthogonal variation in a domain that is not archetypal (i.e., visual processing for typical AD, general cognitive status for PCA).

General cognitive status

In typical AD and PCA there is a considerable amount of variance that is attributable to general cognitive status. However, only by situating PCA within the AD principal component ‘space’ was it possible to compare the severity of general cognitive impairment across these conditions. By plotting individuals from both groups along the AD general cognitive status component we found that the entire range of variation was sampled fully by both groups; in other words, general cognitive status was completely overlapping in typical AD and PCA and there was no evidence of a categorical boundary between them. Even the lower end of the general cognitive status dimension was occupied by both typical AD and PCA. This is
noteworthy as the typical AD cases might have uniquely occupied this region of the multidimensional space given that the tests which comprise the general cognitive status components include measures which are, by definition, fundamentally affected in typical AD, e.g., short-term recognition memory. Instead, this result shows that general cognitive status can be as compromised in PCA as in typical AD, thereby highlighting the importance of non-visual impairments in PCA.

**Visual processing impairments**

Although visual processing impairments are not part of the archetypal presentation of typical AD, our results highlight that such deficits can be detected when they are assessed with a detailed visual battery. We found that cases of typical AD varied gradedly along a dimension representing visual processing impairments, suggesting that variation in visual processing is a feature of phenotypic variation in typical AD. Changes in visual processing are a known symptom of later-stage Alzheimer's disease (Mendez et al., 1990, Kirby et al., 2010), but we ruled out that these visual processing impairments were correlated with general disease severity in our cohort.

The cases with typical AD that we found to have impaired visual processing relative to their general cognitive status performance represent a PCA-like atypical presentation within typical AD. This finding is in agreement with others who have reported cases of typical AD with deviant visual processing performance (Lampon Ralph et al., 2003, Kanne et al., 1998, Martin et al., 1986, Fisher et al., 1999). In our study, these PCA-like typical AD cases occupied the bottom right quadrant of the AD ‘space’. When plotting individuals with PCA into the AD multidimensional space, there was complete overlap between PCA-like typical AD and PCA especially in this lower right quadrant. As expected, only PCA cases occupied the extreme negative end of the AD-visual processing component, which is in line with their cardinal visual symptoms. This shows the variation in severity of visual processing impairments in PCA, and supports the hypothesis that PCA-like typical AD cases represent the mild end of a spectrum of visual processing.

Two components capturing variation in visual processing were extracted for PCA, representing visuoperceptual processing and visuospatial processing. When plotting PCA cases in the space defined by these components, there were graded differences along the visuoperceptual and visuospatial dimensions. This finding is in line with previous research which has found overlapping behavioural presentations (Tsai et al., 2011, Tang-Wai et al., 2004, Mendez et al., 2002), and overlapping distributions of atrophy in the proposed dorsal/ventral subtypes of PCA (Migliaccio et al., 2012, Lehmann et al., 2011a). There were individual cases who scored high on one visual dimension and low on the other (top left and bottom right of the scatter plots in Figure 5.1), implying a selective impairment in visuoperceptual or visuospatial processing. However, there was also no evidence of distinct groups of cases displaying these selective presentations. Thus, studies which seek to compare presentations of PCA by defining groups using inclusion criteria, cut-offs (e.g.,
Lehmann et al., 2011a) or clustering methods (e.g., Ridgway et al., 2012, Tsai et al., 2011) might give the impression of categorical distinctions between these forms of PCA by only sampling/comparing these corners of the phenotypic ‘space’. Overall, our finding of graded differences in visual processing in PCA support the hypothesis that variation in this form of atypical AD occurs along two phenotypic spectra, namely visuospatial processing and visuoperceptual processing (Crutch et al., 2017).

**Conclusions**

By situating the cognitive and visual profiles of impairment in typical AD and PCA in the same multidimensional space, this study elucidated the graded differences between, and heterogeneity within, these forms of AD. An important clinical implication of this result is that considering phenotype variations along multiple, continuous dimensions (i.e., a transdiagnostic approach) might reveal clearer relationships between phenotype and atrophy patterns or genetic markers. For example, past work in post-stroke aphasia (a similarly heterogeneous neurological condition that is also subdivided into diagnostic subtypes) has shown that using continuous dimensions of behavioural performance derived by principal component analysis in voxel-based morphometry analyses reveals unique neural correlates of language symptoms (Butler et al., 2014, Halai et al., 2017). This multidimensional, transdiagnostic approach has recently been applied to the heterogeneous conditions caused by fronto-temporal lobar degeneration (FTLD). Murley et al. (2019) assessed the presence/absence of clinical diagnostic symptoms across in behavioural variant fronto-temporal dementia, non-fluent variant PPA, semantic variant PPA, logopenic variant PPA, progressive supranuclear palsy, and corticobasal syndrome. They established co-varying relationships between shared symptom dimensions and regions of brain atrophy. Importantly, these brain-behaviour relationships revealed graded variation amongst their highly diverse cohort, rather than reproducing the categorical differences between diagnostic classifications. Thus, considering graded differences across diagnostic groups can highlight unique relationships between cognitive symptoms and the extent of underlying brain injury. This multidimensional approach with principal component analysis and voxel-based morphometry could be applied to the heterogeneity in AD to explore the network degeneration hypothesis (Warren et al., 2012) by situating the brain-behaviour relationships of different forms of AD in the same multidimensional symptom-atrophy space. This would build on the important work presented by Seeley et al. (2009) by capturing graded differences between subgroups of neurodegenerative disease instead of comparing groups of cases based on their diagnostic label.

Exploring continuous variation using principal component analysis could also investigate the concept of phenotypic spectra in other neurodegeneration conditions. For example, the concept of a phenotypic spectrum has also been proposed for progressive non-fluent aphasia (Knibb et al., 2009) based on continuous variation in characteristics of connected speech in a cohort of heterogeneous cases of PNFA. Furthermore, although our study only
compared typical AD and PCA, the phenotypic spectrum of AD has been proposed to also include the frontal and logopenic presentations (Ridgway et al., 2012, Migliaccio et al., 2009, Crutch et al., 2012b).

Taking this multidimensional approach could inform a transdiagnostic selection process for treatment, therapy or clinical trials; irrespective of their clinical diagnosis, patients with relatively homogeneous behavioural symptoms could be selected based on occupying a shared region of the multidimensional space (thereby sharing symptomatology across the cognitive systems captured by the dimensions). The importance of a transdiagnostic approach has been highlighted for fronto-temporal lobar degeneration with regards to shared apathy and impulsivity symptomatology across the various conditions which share this pathological substrate (Lansdall et al., 2017, Passamonti et al., 2018, Lansdall et al., 2019). Uncovering shared dimensions of variance across seemingly disparate presentations of AD could aid understanding of common symptoms and therefore development of symptomatic treatments (Murley et al., 2019, Passamonti et al., 2018).

In conclusion, this study represents the first application of a graded, multidimensional framework to situate the cognitive and visual impairments in typical AD and PCA in the same phenotypic space. We found graded variation in visuoperceptual and visuospatial impairments in PCA, rather than evidence of distinct subtypes with selective impairments these visual domains. Furthermore, we found a spectrum of visual processing impairments between the typical AD, PCA-like AD, mild PCA and more severe PCA. Our results therefore support the conceptualisation of (a) two phenotypic spectra in PCA rather than discrete visuoperceptual/visuospatial subtypes, and (b) a spectrum of visual impairments in typical and atypical AD.
Chapter 6 - Discussion

This final chapter is split into eight sections. Firstly, I will summarise the key findings of each empirical chapter. Secondly, I will explore the findings of Chapters 2 and 4 in relation to the ‘primary systems’ hypothesis of language representation in the brain. Thirdly, I will discuss the findings from Chapters 2, 4 and 5, of graded intragroup and intergroup differences in aphasia and also in Alzheimer’s disease. Then, two emergent themes which arose from the direct comparisons of PSA and PPA will also be discussed in the context of relevant literature, namely the uniqueness of semantic dementia (SD), and the importance of assessing beyond impairments in expected domains (i.e., beyond language impairment in aphasia). I will then discuss limitations of the approaches employed in the thesis, and finally place the thesis findings in the context of (a) traditional approaches in cognitive neuropsychology, (b) a vision for clinical applications, and (c) future avenues of research.

Summary of findings

A core aim of the thesis was to compare directly the full ranges of PSA and PPA using a graded, multidimensional framework to map out graded intragroup and intergroup differences. Previous comparisons of PSA and PPA have often been limited in the aphasic presentations or linguistic/cognitive symptoms which were assessed. For example, many have compared PSA and PPA on a specific language/cognitive task (Budd et al., 2010, Thompson et al., 2012, Thompson et al., 2013, Faria et al., 2013). Other studies have focused on comparing ‘fluent’ (Silveri et al., 2019), or ‘non-fluent’ presentations (Patterson et al., 2006a), or those with semantic deficits (Ogar et al., 2011, Jefferies and Lambon Ralph, 2006). A drawback of these comparisons is that selecting presentations of PSA and PPA based on membership of a diagnostic category assumes that the subtypes can be readily identified and are the most appropriate basis for the comparison. However, there is evidence to suggest that subtypes of PSA and PPA are not always clearly differentiable. In fact, there is significant variation within the proposed subtypes and fuzzy boundaries between subtypes, resulting in a considerable proportion of patients who must be classified as having ‘mixed’ aphasia because they either do not fulfil the criteria for any subtype, or even fulfil the criteria for more than one subtype (Benson, 1979; Botha et al., 2015; Gil-Navarro et al., 2013; Harris et al., 2013; Knibb et al., 2009; Matias-Guiu et al., 2014; Mesulam et al., 2008; Mesulam et al., 2012; Sajjadi et al., 2012a; Spinelli et al., 2017; Utianski et al., 2019; Wertz et al., 1984; Wicklund et al., 2014).

If diagnostic categories are not the most appropriate basis for capturing variation in, and making comparisons of, PSA and PPA, then what is? Given that there is clearly an underlying structure to variation in aphasia (i.e., heterogeneity in aphasia phenotype is not caused by random variation and noise), if it is not categorical then it may be continuous. Specifically, variation in aphasia may reflect graded variations along different dimensions,
rather than distinct categories of person/patient. This alternative conceptualisation of patient variance along multiple continuous dimensions may be more in line with the sophisticated way that diagnostic labels are applied in contemporary clinical practice for PSA; the labels provide an indication of the heterogenous constellation of strengths and weaknesses making up a particular person's aphasic profile, rather than to imply that the patient fits the exact diagnostic criteria for a specific aphasia subtype.

The motivation behind reconceptualising patient variance along multiple graded dimensions comes from recent approaches which have employed a non-categorical approach to exploring variation in PSA (Butler et al., 2014, Mirman et al., 2015a). This approach uses principal component analysis (PCA) to extract the core underlying dimensions of variance in aphasia across proposed diagnostic subtypes, i.e., transdiagnostically. Using this framework, graded differences within and between subtypes of PSA have been demonstrated (Butler et al., 2014, Halai et al., 2017, Schumacher et al., 2019). Furthermore, this graded, multidimensional approach has enabled the extraction of unique neural correlates associated with the principal dimensions of aphasia in PSA.

**Chapter 2** represents the extension of this multidimensional framework in two important directions. Firstly, this study is the first to apply PCA to PPA in order to explore the graded differences between proposed subtypes. Secondly, this study is the first to use the PCA framework as a platform for comparing PSA and PPA by extracting and characterising the shared multidimensional space of these forms of aphasia. The results of Chapter 2 showed that the principal dimensions of variance in PSA and PPA were remarkably similar, comprising variance in phonology, semantics, visuo-executive function and motor speech production. Furthermore, creating a shared test battery enabled the extraction of shared dimensions of variance. PSA and PPA were then plotted in this shared space to reveal graded intra-group and intergroup differences. The exception was semantic dementia (SD) which occupied a unique part of the multidimensional space, reflecting the selective and substantial semantic impairment in this group which is not present in any other presentation of PSA and PPA. There was also a key difference between PSA and PPA in terms of speech output fluency; plotting all cases in the transdiagnostic shared multidimensional space revealed that ‘fluent’ PSA often had lower scores on the ‘Motor speech production’ dimension than ‘non-fluent’ PPA.

The creation of a shared test battery for PSA and PPA was fundamental to the work in Chapter 2. The lack of a standardised linguistic tool which has been validated for aphasia across these aetiologies has likely contributed to the paucity of systematic comparisons of PSA and PPA. The Mini-Linguistic State Examination (MLSE) is a novel clinical assessment tool which is in development to address this gap. **Chapter 3** completed the first formal validation of the MLSE and compared it to established aphasia tests. This study compared the performance of people with PSA on the established tests and the MLSE for two key reasons: (1) the established aphasia tests have been designed and validated for PSA, and
(2) it enabled us to establish the profiles of MLSE performance in mild and moderate-to-severe PSA subtypes, laying the groundwork for future comparisons of PSA and PPA using this novel tool. The validity of the MLSE was assessed in terms of: (1) convergence of the subtype classifications derived from MLSE test scores vs. test scores on the Boston Diagnostic Aphasia Examination; (2) sensitivity of the MLSE compared to a test battery of more in-depth tests in terms of (a) sensitivity to mild deficits, and (b) sensitivity to levels of impairment; and (3) convergent validity of the MLSE subtests compared with the in-depth tests. Overall, we found that the MLSE is hypersensitive to the presence of a language impairment in PSA and that many of the MLSE subtests show good convergent validity.

**Chapter 4** built on the preceding chapters by utilising the MLSE to explore graded variation in language impairments across a cohort of PSA, PPA and also corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP). The latter two cohorts were included because language impairments are increasingly recognised as a significant symptom in these forms of fronto-temporal lobar degeneration (Peterson et al., 2019) but they have not been systematically investigated in comparison with PSA and PPA. In addition to comparing aphasia across these diverse aetiologies, Chapter 4 utilised structural magnetic resonance imaging (MRI) to explore the neural correlates of these language impairments. The transdiagnostic nature of this study was important because it leveraged the differences between PSA and PPA in order to learn more about aphasia. For example, the regions of the language network that are affected in PSA (e.g., the territory of the middle cerebral artery) only partially overlap with those affected by the neurodegenerative processes in PPA (e.g., bilateral ventrolateral anterior temporal lobes are rarely affected by MCA stroke).

Performance on the MLSE captured characteristic language impairments for each diagnostic group, which were accompanied by typical patterns of neural abnormality. Briefly, the heterogeneous PSA cohort (the same cohort as reported in Chapter 3 - comprising a spectrum of behavioural presentations across different severities) was associated with impairments in all linguistic domains of the MLSE and neural abnormality in a large perisylvian region attributable to the vascular territory of the middle cerebral artery; PNFA was associated with impairments in sentence comprehension and abnormality was found in a left-dominant pattern encompassing posterior temporal cortex and underlying white matter; LPA was associated with impaired sentence repetition and abnormality an asymmetrical, left-dominant pattern encompassing superior and posterior temporal lobe grey and white matter; SD was associated with impairment semantic knowledge and asymmetrical, left-dominant abnormality of the anterior temporal lobes; CBS cases displayed mild deficits in targeted MLSE subtests relative to controls, and showed abnormality in left posterior cingulate cortex, bilateral supplementary motor area, left precentral gyrus, and left occipital pole. PSP was associated with impairments in writing and picture description (and mild deficits in oral reading) and abnormality in brainstem and midbrain regions.
Chapter 4 also applied the graded, multidimensional PCA framework outlined in Chapter 2 to the MLSE dataset; shared dimensions of phonology, semantics, and syntax were found to characterise transdiagnostic variance on the MLSE across aetiologies. These dimensions replicated the findings of Chapter 2, providing further support for the existence of core ‘primary systems’ underlying language impairments in aphasia. Furthermore, Chapter 4 utilised a type of voxel-based morphology analysis called voxel-based correlational methodology (VBCM) to reveal unique neural correlates for the three extracted principal dimensions. The phonology dimension was uniquely related to neural abnormalities in left hemisphere temporoparietal regions including angular gyrus, supramarginal gyrus, insula, and white matter tracts of the ventral language route. This result is consistent with previous VBCM analyses in PSA which have also shown that phonological processing abilities are related to the supramarginal gyrus (Sliwinska et al., 2012, Stoeckel et al., 2009, Price et al., 1997, Newman and Twieg, 2001, Dehaene-Lambertz et al., 2005), and planum temporale (Griffiths and Warren, 2002, Jäncke and Shah, 2002). The semantic dimension was uniquely related to neural abnormalities in bilateral anterior temporal lobes, which is in line with previous findings from transcranial magnetic stimulation (Pobric et al., 2007), intraoperative stimulation studies (Duffau et al., 2008), imaging studies of SD (Mion et al., 2010), functional MRI (Vigneau et al., 2006, Visser et al., 2010, Binder et al., 2009), and computational models (Ueno et al., 2011). The syntax dimension was uniquely related to neural abnormalities in left lateral ventricle, left caudate, and nearby temporal lobe white matter tracts of the ventral language route (IFOF and ILF) (Parker et al., 2005, Bajada et al., 2015). This cluster encompassed regions which have been associated with functions including speech generation (Crosson et al., 2003, Crinion et al., 2006), and reading and writing (Tomasino et al., 2015). Thus, by comparing linguistic impairments on the same test battery, in the same multidimensional space, and using the same neuroimaging approach, Chapter 4 uncovered brain-behaviour relationships that underly variance in aphasia across aetiologies.

Chapter 5, the final empirical chapter in the thesis, represents an extension of this graded, multidimensional PCA approach to variation in Alzheimer’s disease (AD). As with aphasia, a shift in the zeitgeist of research into AD is also occurring in the direction of reconceptualising patient variance in terms of multiple phenotypic spectra. Like presentations of PSA and PPA, AD is also subdivided into typical and atypical subtypes in recent diagnostic consensus recommendations (Dubois et al., 2014). Whether these subtypes (or variants) represent categorically distinct subtypes of AD or whether they would be better conceptualised as varying from each other on one or more phenotypic spectra remains to be established. Atypical AD variants include frontal-variant AD, language-variant AD (also known as logopenic progressive aphasia), and visual-variant AD (also known as posterior cortical atrophy). Furthermore, posterior cortical atrophy has been proposed to comprise two subtypes; visuospatial variant posterior cortical atrophy is defined as having a disproportionate impairment in the dorsal (‘Where’) visual processing stream. Visuoperceptual variance posterior cortical atrophy is instead defined as a relatively greater
impairment in the ventral (‘What’) visual processing stream. However, as with the relationship between typical and atypical variants of AD, whether these proposed subtypes of posterior cortical atrophy are categorically distinct has yet to be explored. To address this question, Chapter 5 applied the graded, multidimensional PCA framework to Alzheimer’s disease and posterior cortical atrophy. A large database collected from people with typical (i.e., amnestic) AD and posterior cortical atrophy on a variety of cognitive and visual tests was analysed using PCA. For typical AD, the PCA extracted two components reflecting general cognitive status and visual processing. Plotting the cases with posterior cortical atrophy along the dimensions from the AD PCA highlighted a continuum in visual processing ability. Cases with typical AD uniquely occupied the ‘top’ of this dimension, reflecting spared visual processing ability. Conversely, cases with posterior cortical atrophy uniquely occupied the ‘bottom’ of this dimension, reflecting globally impaired visual processing ability. However, between these extremes, cases of typical AD were found to have visual processing impairments that overlapped with cases of mild posterior cortical atrophy. This result highlights (a) the importance of measuring more than memory in typical AD, (b) the utility of the PCA method for visualising transdiagnostic variation, (c) the heterogeneity within the proposed typical variant of AD, and (d) the lack of a categorical distinction between typical AD and posterior cortical atrophy. For posterior cortical atrophy, the PCA extracted three components reflecting general cognitive status, visuoperceptual processing, and visuospatial processing. At first glance this result appears to confirm the existence of separable dimensions of variance in the visuoperceptual (‘What’) symptoms and visuospatial (‘Where’) symptoms in posterior cortical atrophy. However, plotting all posterior cortical atrophy cases along these visual dimensions demonstrated considerable graded variation along both dimensions, contradicting the idea of categorically different presentations of posterior cortical atrophy. The results of Chapter 5 show the generalisability and utility of the graded, multidimensional PCA framework employed in Chapters 2-4 in aphasia to another neurological condition.

Uncovering the primary systems of language

Chapters 2 and 4 involved systematic comparisons of the full ranges of PSA and PPA using a multidimensional approach to map out the graded intragroup and intergroup differences in language impairments and their neural correlates. In Chapter 2 we uncovered components of aphasia which appear to correspond to primary systems of language transdiagnostically (i.e., not aetiology specific), namely primary systems for phonology, semantics, and speech production. In Chapter 4 we replicated these components and uncovered the neural correlates of these dimensions. Table 6.1 summarises the principal dimensions extracted for different cohorts in Chapters 2 and 4.

The nature of the underlying dimensions of variance in aphasia remained remarkably stable when considering (a) PSA or PPA separately (Chapter 2 separate PCAs for PSA and PPA), (b) aetiology-specific vs. shared multidimensional spaces (Chapter 2 separate PCAs vs.
shared PCA), and (c) different cohorts and test batteries (Chapter 2 shared PCA vs. Chapter 4 PCA). Figure 6.1 shows the commonalities amongst PCA results in the thesis. The fact that consistent underlying dimensions for phonology, semantics, and connected speech/syntax were found across PSA and PPA indicates that these dimensions might reflect core “primary systems” for language activities (Madden et al., 2018, Savill et al., 2019, Patterson and Lambon Ralph, 1999, Ueno et al., 2014, Woollams et al., 2018). The primary systems hypothesis proposes that there are domain-general systems which support functions across task demands (Patterson and Lambon Ralph, 1999, Plaut et al., 1996, Seidenberg and McClelland, 1989). The primary systems hypothesis arose through observations of SD (Patterson et al., 2006b), and has found support in evidence focused on healthy participants (Savill et al., 2019), and PSA (Woollams et al., 2018, Madden et al., 2018, Ueno et al., 2014, Crisp and Lambon Ralph, 2006, Lambon Ralph et al., 2002).

Woollams et al. (2018) tested the primary systems hypothesis by relating reading deficits to underlying neural correlates in PSA; in their study, they applied PCA to performance on an extensive neuropsychological test battery and found that three components, termed phonology, semantics, and cognition, comprised the dimensions of the phenotypic space of their sample. They related the phonology and semantics components, as well as components from PCA on a battery of reading measures, to neural correlates using VBCM. Their phonology principal component was associated with neural correlates in a large left hemisphere region encompassing frontal pole, middle and inferior frontal gyri, posterior insula, planum polare and planum temporale, superior and middle temporal gyri, and posterior supramarginal gyrus. Their semantics principal component was associated with neural correlates in the left hemisphere, including anterior temporal lobe, temporal pole, planum polare, middle and inferior temporal gyri, anterior fusiform cortex, and the inferior longitudinal fasciculus. Woollams et al. (2018) confirmed that these clusters encompassed domain-general regions which supported performance across different linguistic processes by overlapping the proposed primary systems clusters with clusters associated with different reading tasks. The authors found that the phonology cluster overlapped with clusters associated with varied reading tasks, including non-words and words. The semantics cluster was found to overlap with the cluster for word reading but not the cluster for non-word reading. This supports the primary systems hypothesis that the phonology and semantics domain-general systems work dynamically to support different tasks. Therefore, to suggest that the studies in the thesis have captured primary systems in PSA and PPA, the neural correlates in Chapter 4 should have captured the same areas as Woollams et al. (2018).
Table 6.1 – Shared principal dimensions across PSA and PPA.

<table>
<thead>
<tr>
<th>Population</th>
<th>Principal components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSA</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>PPA</td>
</tr>
<tr>
<td>PSA and PPA</td>
<td>PSA, PPA and movement disorders</td>
</tr>
<tr>
<td>Speech generation</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.1 – Commonalities among principal component analyses in the thesis.

This figure shows the factor loadings of similar tests across the PCAs of Chapters 2 and 4. The test labels on the y axis correspond to MLSE subtests from Chapter 4. Factor loadings from tests entered in the PCAs in Chapter 2 were matched to the subtests of the MLSE as closely as possible (the same process as the creation of the Matched battery). Some MLSE subtests did not have a corresponding test in Chapter 2.

Abbreviations: PSA – post-stroke aphasia; PPA – primary progressive aphasia; MD – movement disorders; MLSE – Mini-linguistic state examination.
We found several of the same regions to be associated with our phonology component, including posterior insula, posterior supramarginal gyrus, and planum temporale. Previous research has also found that a PCA component capturing performance on similar tasks (e.g., repetition), and similarly denoted as capturing ‘phonology’, has also been associated with unique neural correlates in posterior insula and planum temporale (Halai et al., 2017, Schumacher et al., 2019). These regions have been proposed to be processing ‘hubs’ (Griffiths and Warren, 2002) or to provide connections between diverse language-processing areas (Ardila et al., 2014, Oh et al., 2014), and thus appear to be appropriate candidates for the primary system of phonology. Thus, the VBCM analysis in Chapter 4 appears to have replicated the neural correlates of the primary system of phonology simultaneously in a cohort of PSA and PPA. As these perisylvian regions can be affected in both middle cerebral artery stroke (Phan et al., 2005) and PPA (Gorno-Tempini et al., 2004), the overlap in their phonological deficits could reflect varying degrees of impairment to this primary system.

However, the phonology cluster in Chapter 4 did not encompass several regions which were found by Woollams et al. (2018). For example, neural abnormality in the superior temporal gyrus and middle temporal gyrus was not associated with worse performance on our phonology component. These regions have previously been associated with a phonology component by other groups (Halai et al., 2017, Schumacher et al., 2019), and therefore not finding them is unexpected. A possible reason for this is that the test batteries employed by the aforementioned studies contained more tests of receptive phonology, compared to the MLSE in Chapter 4 which is more focused on expressive phonology. Expressive and receptive phonological processing have been shown to have separable neural correlates in more posterior and anterior locations, respectively (Alyahya et al., in press). This could explain the more posterior location of the phonology cluster in Chapter 4, compared to previous studies with more anterior clusters encompassing the superior and middle temporal gyri (Woollams et al., 2018, Halai et al., 2017, Butler et al., 2014, Schumacher et al., 2019). Therefore, the results of Chapter 4 may only provide evidence of the more posterior aspects of the phonology primary system, and further work is needed to explore this system in PSA and PPA more fully.

The semantic cluster in Chapter 4 encompassed several regions which were also part of the semantics cluster in the study by Woollams et al. (2018). These included the anterior temporal lobe and temporal pole, inferior temporal gyrus, and anterior fusiform cortex. In Woollams et al. (2018) these regions were part of a unilateral cluster in the left hemisphere (reflecting the unilateral nature of the left-hemisphere stroke aetiology in their cohort), but in Chapter 4 we found that the semantic component was associated with bilateral neural abnormality in these regions (reflecting the bilateral nature of neurodegeneration in the PPA and movement disorder cohorts).

The ATL has been robustly associated with a principal dimension capturing semantic impairments in previous studies (Butler et al., 2014, Halai et al., 2017, Schumacher et al.,
The left and right anterior fusiform cortex have been linked to verbal and non-verbal semantic tasks, respectively (Mion 2010). As expected given the similar reading tasks in Woollams et al. (2018) and in the MLSE, we also found that the posterior inferior temporal gyrus was encompassed by the semantic cluster, as this region has been associated with reading of words vs. non-words (Fiez et al., 1999; Hagoort et al., 1999), and semantic association (Booth et al., 2002). However, Woollams et al. (2018) found neural correlates for their semantic component in regions which we did not find in Chapter 4. In particular, we did not find an association between poor performance on our semantic component and neural abnormality in the left inferior longitudinal fasciculus (ILF). Instead, we found an association with the right inferior fronto-occipital fasciculus (IFOF). The ILF is an associative, long-range white matter tract which connects posterior, occipital regions of the brain to anterior, temporal regions; this tract is directly connected to the IFOF but is situated more ventrally and superficially (Herbet 2018). Both the ILF and IFOF have been related to the semantic ventral stream and implicated in semantic processing (Duffau et al., 2013). Agosta et al. (2010) showed that the ILF is impacted in SD, so it is unexpected that we did not detect an association with poor semantic performance and neural abnormality in this region. However, the IFOF has been implicated in PSP, so the inclusion of the PSP cohort in Chapter 4 could have driven the association with the IFOF instead (Kvickström et al., 2011). Further research is needed to establish the status of potential domain-general systems supporting semantic cognition in PSP and other movement disorders, which possibly show milder semantic impairments than occur in aphasia caused by stroke or neurodegeneration.

Overall, the neural correlates for semantics in Chapter 4 converge with Woollams et al. (2018) and previous findings from transcranial magnetic stimulation (Pobric et al., 2007), intraoperative stimulation studies (Duffau et al., 2008), imaging studies of SD (Mion et al., 2010), functional MRI (Vigneau et al., 2006, Visser et al., 2010, Binder et al., 2009), and computational models (Ueno et al., 2011), showing the importance of the anterior temporal lobes, fusiform cortex, and associated white matter for semantic cognition across different task demands. Thus, the results of Chapter 4 appear to have replicated the neural correlates of the semantics primary system and extended this to bilateral neural correlates, building on Woollams et al. (2018) with the addition of information from non-stroke aetiologies.

It is less clear whether the results of Chapter 4, and other PCAs across the thesis, support a third, speech production component of the primary systems framework. Woollams et al. (2018) found a cluster in the left hemisphere, in a superior frontal region which was associated specifically with non-word reading. The regions encompassed by this cluster included inferior frontal gyrus, pre- and post-central gyrus, opercular cortex, and the arcuate fasciculus. Previous groups have found these regions to be associated with speech quanta (Catani et al., 2013) or motor speech output (Price, 2012). Woollams et al. (2018) therefore added a component to the primary systems framework for motor speech output, supported
by this frontal region. Halai et al. (2017) also found a neural correlate for a component reflecting speech quanta, which encompassed the precentral gyrus, putamen and insula, thereby appearing similar to the ‘fluency’ cluster found by Woollams et al. (2018).

The syntax component-cluster in Chapter 4 does not seem to reflect these speech production components or the same neural correlates. The component contained loadings from tests of production and comprehension of grammatically complex sentences, whilst the neural correlates included the left lateral ventricle, left caudate, and nearby temporal lobe white matter tracts. The loadings from measures of receptive syntactic processing (e.g., sentence comprehension without verbal response) make this component different from the previous sentence production components described above. Although perhaps not capturing the motor speech production primary system, the neural correlates of the syntax component have been associated with linguistic processes which are involved in the tasks that loaded heavily onto this component. For example, the cluster encompassed the IFOF and ILF, which are part of the ventral language route (Parker et al., 2005, Bajada et al., 2015) and damage to the IFOF has been associated with reading and writing impairments (Tomasino et al., 2015). In general, the PSP and CBS were relatively spared on the phonology and semantics dimensions but displayed significant impairments on sentence-level tests (e.g., sentence comprehension and writing). As such, we could speculate that this syntax component-cluster may be driven by the movement disorders, but this needs further exploration in future studies. Interestingly, the syntax component did also have a strong loading from non-word reading. Thus, it seems possible that the verbal expressive tasks loading onto this component could reflect motor speech production. With more data, it may have been possible to extract a fourth factor, which could result in the expressive/receptive aspects of sentence processing being split.

**Graded intragroup and intergroup differences**

Chapters 2, 4 and 5 highlighted considerable overlap in behavioural deficits across diagnostic groups. Having highlighted dimensions of variance which are consistent across diagnostic categories (e.g., phonology and semantics in aphasia, visual processing in typical AD and posterior cortical atrophy), a subsequent question that arises is whether these diagnostic groups have quantitatively overlapping deficits on these dimensions.

**Intergroup, transdiagnostic variation**

In Chapter 2 and Chapter 4, there was considerable overlap in the range of phonology impairments between PSA and non-SD PSA; SD were, as expected, relatively spared in this domain. In Chapter 4, the CBS and PSP cases were also relatively spared in terms of phonology. This transdiagnostic separation might reflect differing patterns of brain injury to perisylvian (PSA and non-SD PPA) vs. extrasylvian (SD, CBS and PSP) language regions. PSP is characterised by atrophy in the midbrain (Soliveri et al., 1999, Gröschel et al., 2004).
SD arises from atrophy in extra-sylvian, ATL regions (Snowden et al., 1989, Hodges et al., 1992, Rosen et al., 2002, Mummery et al., 2000, Mion et al., 2010). Atrophy in CBS is found in perisylvian language regions such as the insula, but also in extrasylvian regions such as premotor cortex and supplementary motor area (Whitwell et al., 2010) and deep-brain subcortical structures such as striatum (Boxer et al., 2006). In contrast, middle cerebral artery stroke primarily affects cortical and subcortical perisylvian language regions (Phan et al., 2005, Hillis et al., 2002, Hillis et al., 2004), and PNFA and LPA are also associated with damage to perisylvian cortical and subcortical regions (Grossman and Irwin, 2018, Gorno-Tempini et al., 2004, Gorno-Tempini et al., 2011).

In terms of semantics, SD selectively occupy the negative end of only the semantic dimension across PCAs in Chapters 2 and 4. The uniqueness of SD in aphasia is discussed further below. Some cases of PSA and non-SD PPA occupy a similar negative region of the semantic dimension, but this is tends to be concurrent with occupying the negative end of the other dimensions (e.g., bottom left quadrant of scatter plots in Figure 2.2 panel A, and Figure 4.2, panels A and C). Thus, severe semantic impairments in PSA and non-SD PPA co-occur with severe impairments in other linguistic domains. This could possibly reflect (a) more severe aphasia in these cases due to a larger stroke lesion (Thye and Mirman, 2018) or further progression of atrophy due to later disease stage, such that regions supporting semantic impairment are directly compromised (Rohrer et al., 2013, Rogalski et al., 2011), or (b) severe deficits in other domains (e.g., speech output or executive processing (Keil and Kaszniak, 2002)) confounding the assessment of semantic impairments.

Chapter 2 revealed important variation in visuo-executive function in PSA and PPA. Furthermore, there was a possibly aetiology-driven split whereby cases with PSA did not tend to vary on the visuo-executive factor. Cases of PNFA, LPA and mixed PPA did vary considerably on this factor, which is in line with previous research showing that visuospatial deficits are a key part of the constellation of impairments in PPA (Watson et al., 2018). The PSA cases were relatively spared on the visuo-executive factor, which could be due to the extent of their lesions being dictated by the middle cerebral artery (MCA) territory. For example, the multi-demand frontoparietal executive system (Marek and Dosenbach, 2018), and posterior cingulate and other medial regions which support executive function and attention (Jurado and Rosselli, 2007) are situated at the edges/outside of the MCA-perfused regions (Phan et al., 2005). This might mean that the cognitive functions supported by regions at the edges of/outside the territory of the MCA would be less likely to be impaired in PSA than perhaps in some forms of PPA.

In terms of the connected speech components in Chapters 2 and 4, there was a general separation between PSA and non-fluent PPA vs. fluent PPA (and movement disorders). In Chapter 2, poorest speech production performance was found in PSA and some mixed PPA cases. Overall though, PNFA cases tended to have better speech production performance than even the Anomic PSA cases. However, in Chapter 4, where the speech component
seemed to reflect syntax/comprehension as well as production, the PNFA and PSA cases were highly overlapping. This suggests that it is the expressive/motor aspect of speech production which might differentiate fluency in PSA and PPA. Thompson et al. (2013) found that cases of PSA and PPA with agrammatism had similar impairments in syntactic and morphosyntactic processing. However, Karbe et al. (1993) found that patients with PPA were more impaired on the WAB spontaneous speech measure than PSA and AD; specifically, patients with PPA were worse on the fluency scores than both PSA and AD, but only worse for the information content measure compared to AD. This supports the thesis findings that PSA and PPA have overlapping connected speech impairments in terms of syntax/content but not speech production/fluency. This result is also consistent with previous direct comparisons restricted to the non-fluent subtypes of PSA and PPA which showed that despite similar nomenclature, they do not display equivalent levels of fluency (Patterson et al., 2006a). This separation in fluency across PSA and PPA is clinically interesting and important as it indicates that certain symptom terms – e.g., fluency – are not used in the same way across patient types; thus, many non-fluent progressive aphasics were more fluent than the “fluent” PSA cases (e.g., anomic and conduction aphasics). This may be relevant for clinical professionals who work with people with PSA and with people with PPA; if assessments/tools at their disposal are targeted towards ‘non-fluent’ aphasias then it may be useful to have a formal overview of how the term ‘fluency’ is applied across the shared space of PSA and PPA.

As discussed above and in Chapters 2 and 4, differences in the location of neural damage could be responsible for intergroup variation in phonological and semantic impairments in PSA and PPA. However, as well as affecting different neuroanatomical regions, stroke and neurodegeneration cause neural damage through different cellular mechanisms and with different time courses. Therefore, the nature and timing of the neural damage could also contribute to intergroup differences between PSA and PPA.

For example, having time to adapt to the slow loss of function in a degenerating brain region may allow other regions to adopt the function and so minimise any detectable impairment. This was demonstrated by Keidel et al. (2010) who compared aphasia due to slow-growing glioma vs. post-stroke aphasia. These authors found that one of the important factors behind milder aphasia in low grade glioma patients compared to PSA patients was the difference in timing of the damage. Using computational modelling to support observations in patient profiles, they proposed that representations stored in affected tissue can guide the restructuring of new cortical networks to take on the at-risk function, if the affected representations is at least partially available. Thus, in low grade glioma, new cortical structures can adopt the function of the affected tissue because it degenerates slowly. Hence, the affected tissue remains available to guide the development of representations in the new structures, meaning language-processing functions can be largely maintained. On the other hand, the representations stored in the affected tissue in stroke are made
completely unavailable for guiding the restructuring of new cortical regions, resulting in worse aphasia. The different effects of the timing of damage in stroke and glioma were also demonstrated by Thiel et al. (2005) who showed that right hemisphere language-processing regions were able to take on the usually left-lateralised role of left hemisphere regions affected by glioma only if the glioma was slow growing. Thus, the timing of neuronal injury has important implications for the profiles of language impairment in PSA and PPA.

One difference in the nature of damage in stroke compared to neurodegeneration is the laterality of the injury. PSA is usually caused by unilateral stroke in the left hemisphere (Berthier, 2005), whilst neurodegeneration in PPA is often bilateral (e.g., in semantic dementia (Acosta-Cabronero et al., 2011) and late-stage logopenic progressive aphasia (Rohrer et al., 2013)). Schapiro et al. (2013) showed that unilateral damage reduces the magnitude of contribution from one hemisphere but the other hemisphere can overcome this by becoming dominant; in contrast, bilateral damage distorts and reduces contributions from both hemispheres, meaning overall performance is not maintained. Thus, bilateral damage in PPA may be more likely to contribute to overall worse performance than PSA due to unilateral stroke.

Another difference in the nature of damage due to stroke or neurodegeneration is the extent to which white matter and subcortical structures are also affected. In stroke, which regions and cell types are vulnerable is partially determined by the amount of nearby microvasculature (Nasrabady et al., 2018, Nonaka et al., 2003). This means that deep white matter is particularly vulnerable and often accounts for up to 50% of the infarct volume (Ho et al., 2005). In contrast, grey matter at specific cortical layers is typical of neurodegenerative disease (Romito-DiGiacomo et al., 2007). For example, degeneration is often seen in layers II (Gómez-Isla et al., 1996) and IV (Hyman et al., 1984) in Alzheimer’s disease, and layer II in fronto-temporal lobar degeneration (Mackenzie et al., 2006). White matter changes in neurodegenerative disease are being increasingly recognised (Seeley et al., 2009), including in PPA (Galantucci et al., 2011, Mahoney et al., 2013, Schwindt et al., 2013) and Alzheimer's disease (Nasrabady et al., 2018). However, whether white matter injury in stroke and neurodegenerative disease have equivalent consequences is still unclear. Garcia-Cordero et al. (2015) found that insula cortex damage after stroke and in behavioural variant fronto-temporal dementia resulted in different effects on local connectivity with this region; ischaemic stroke resulted in hypoconnectivity of local voxels in the insula cortex, whereas fronto-temporal dementia pathological degeneration of the insula cortex was associated with hyperconnectivity in local voxels. However, their study was not able to investigate the reason behind these different effects on local connectivity. Overall, further research is needed to compare the mechanism by which stroke or neurodegenerative disease affects different cortical and subcortical structures and connections between them. This could help explain the graded intergroup variation found between some subtypes of
PSA and PPA by elucidating which language-processing structures are affected in similar or different ways by these forms of neural injury.

In Chapter 5, the nature and timing of neural injury were (presumably) consistent across groups (due to the same suspected underlying Alzheimer's disease pathology), but the location of injury likely contributed to the observed graded intergroup differences. Some cases of typical AD showed a relatively selective impairment on the visual processing dimension compared to their scores on the general cognitive status component (occupying the bottom right quadrant of Figure 5.2). These cases therefore echo the profile of visual impairments in posterior cortical atrophy, signifying an atypical-like pattern within typical AD. When posterior cortical atrophy cases were projected into the AD multidimensional space, there was complete overlap with atypical-like AD cases in the region capturing less severe impairments in the visual processing dimension. Thus, the visual impairments in the atypical-like AD cases was equivalent to the milder cases of posterior cortical atrophy. The extreme negative end of the visual processing component was uniquely occupied by posterior cortical atrophy cases, signifying more severe visual impairments in these cases. Thus, we found a spectrum of visual impairments ranging from unimpaired (typical AD), to mildly impaired (atypical-like AD and milder posterior cortical atrophy), to severely impaired (more severe posterior cortical atrophy). This supports the hypothesis proposed recently that typical AD and atypical AD vary along one or more phenotypic spectra (Warren et al., 2012, Stopford et al., 2008, Ridgway et al., 2012, Lambon Ralph et al., 2003, Migliaccio et al., 2009, Fitzpatrick et al., 2019, Peter et al., 2014); we have shown that typical AD and the posterior cortical atrophy atypical variant of AD vary gradedly along a visual processing spectrum.

Intragroup heterogeneity

The PCA method also revealed heterogeneity in linguistic, cognitive, and/or visual symptoms within the diagnostic groups analysed in the thesis studies. In Chapter 2, the PCA revealed heterogenous subtype profiles on the extracted phonology, semantics, visuo-executive and motor speech components. Graded differences between subtypes and within-group heterogeneity were explored formally by classification analysis for each aetiology. Semantic dementia proved to form a true diagnostic category (i.e., within group homogeneity and distinct between group differences), whereas there was considerable graded variation within and between other subtypes of PPA and PSA. These results were replicated in Chapter 4, again showing heterogeneity in the severity of impairments displayed the PSA and non-SD PPA diagnostic groups along the extracted language dimensions. Similarly to SD, the CBS and PSP groups showed less intragroup heterogeneity, especially on the phonology and semantics factors. This could be quantified in future research using classification analysis, as in Chapter 2, to see if data-driven diagnostic cut-offs for CBS and PSP meet the assumptions of a true category.

Within-group heterogeneity was also evidence for typical AD in Chapter 5. The PCA revealed that cases with a diagnosis of typical AD vary along two key behavioural...
dimensions, reflecting general cognitive status and visual processing abilities. From Figure 5.2 it is evident that some typical AD cases displayed a relatively selective impairment on the general cognitive status dimension (top left of the scatterplot in panel A), whilst other typical AD cases showed the opposite pattern (bottom right of the scatterplot in panel A). The latter cases represent a visually-atypical presentation within the diagnostic category of typical AD, which other comparisons of typical AD have also found, along with semantically-atypical or executively-atypical presentations (Lambon Ralph et al., 2003, Kanne et al., 1998, Martin et al., 1986, Fisher et al., 1999).

Finally, heterogeneity within the diagnostic group of posterior cortical atrophy comprised graded variation in visuospatial and visuoperceptual processing. By plotting all posterior cortical atrophy cases along these dimensions, they were scattered evenly across all four quadrants of the space (Figure 5.1), providing no evidence for distinct categories of visual impairments. Some cases had relatively spared visuospatial and visuoperceptual processing (top right; possibly milder cases of the disorder), and some had relatively impaired processing in both visual domains (bottom left; more severe cases). The top left and bottom right corners of this visuoperceptual-visuospatial space signify relatively selective impairments in either visuoperceptual (top left) or visuospatial (bottom right) processing. Crucially, the entire phenotypic space was scattered with cases showing heterogeneity in the extent of impairment on each dimension. Overall, our finding of graded differences in visual processing in PCA supports the hypothesis that variation in this form of atypical AD occurs along two phenotypic spectra, namely visuospatial processing and visuoperceptual processing (Crutch et al., 2017). Comparing only the relatively selectively-impaired cases from the top left/bottom right corners would give the false impression of discrete presentations of visual impairment. Instead, a graded approach such as the one employed here with PCA is required to capture the full phenotypic spectra of visual impairments in posterior cortical atrophy.

**Emergent themes**

Two new themes emerged across the empirical studies of the thesis. Firstly, a variety of extensive and refined test batteries were employed in this thesis, some of which were tailored to assess domains known to be central to the deficits characteristic of the studied disorders. Other test batteries included assessments of cognitive processes that were not necessarily core to the diagnostic categories studied. In combination with the PCA methodology employed across the thesis, the importance of assessing these ‘non-target’ cognitive domains (e.g., executive function in PPA, or visual processing abilities in AD) became increasingly clear. Secondly, SD was consistently found to be unique amongst the diverse aphasic phenotypes studied in the thesis. This was the case when considering the linguistic impairments in PSA and PPA (Chapter 2 & 4), and also the neural correlates of these impairments (Chapter 4). These emergent themes will be discussed further below.
Is semantic dementia unique in aphasia?

Chapters 2 and 4 established the shared multidimensional space of aphasia across subtypes of PSA, PPA, and movement disorders. In every analysis which included semantic dementia, a clear dimension of variance capturing semantic impairments was extracted. From the scatterplots in Chapters 2 and 4 (Figure 2.2 and Figure 4.2), it is clear that SD uniquely occupies the region of multidimensional space signifying selective semantic impairments. This was formally shown in Chapter 2 using the data-driven diagnostic cut-offs which showed perfect selectivity and specificity for SD. By situating different aetiologies of aphasia in the same shared multidimensional space, we found that no other form of aphasia has comparable semantic impairments, either qualitatively or quantitatively. Specifically, in Chapter 2, without SD the nature of the extracted semantic component was different (capturing more variance in comprehension/executive function). In Chapters 2 and 4, the other diagnostic subgroups did not overlap with the region of multidimensional space occupied by SD, nor did they show neural abnormality in the same anterior temporal lobe regions. These results suggest that SD is a unique disorder which involves impairment of the semantic primary system.

Previous work has shown that semantic impairments in SD are unlike those found in PSA or other subtypes of PPA. For example, Jefferies and Lambon Ralph (2006) compared SD to semantically-impaired PSA. They found that semantic impairment across different tasks in SD was attributable to the loss of amodal semantic representations. In contrast, semantic impairment in PSA was attributed to a deficit in accessing intact semantic representations, which could be reversed through cueing (see also Thompson et al., 2015). Furthermore, SD is characterised by atrophy in the ATLs (Lambon Ralph et al., 2017), which are supplied by two arteries (Kiernan, 2012), reducing its vulnerability to damage after stroke and therefore reducing the likelihood of PSA phenotypes which mimic SD.

Previous studies have shown that SD is readily differentiable from the other forms of PPA using cluster analysis (Hoffman et al., 2017), machine learning (Bisenius et al., 2017) and structural MRI atrophy patterns (Sajjadi et al., 2017). The uniqueness of semantic impairment in SD probably reflects the fact that the distribution of damage in SD is distinctly different from atrophy patterns in non-SD PPA, PSA, and other neurodegenerative disorders. As described above, SD arises from atrophy in extra-sylvian, ATL regions (Snowden et al., 1989, Hodges et al., 1992, Rosen et al., 2002, Mummery et al., 2000, Mion et al., 2010). In contrast, PSA, non-SD PPA, and also other presentations of fronto-temporal lobar degeneration (such as CBS and PSP) are characterised by damage to perisylvian cortical and subcortical regions (Grossman and Irwin, 2018, Hillis et al., 2002, Hillis et al., 2004), premotor and supplementary motor cortices (Whitwell et al., 2010), and midbrain regions (Soliveri et al., 1999, Gröschel et al., 2004, Boxer et al., 2006).

The results of the thesis show for the first time that SD is unique amongst other disorders displaying language impairments when considered together transdiagnostically on the same
test battery and using the same neuroimaging measures. Thus, the thesis findings provide support for the concept of ‘SD’ as a uniquely useful pointer for the exclusive region of the aphasic multidimensional space occupied by these cases.

How to capture the full phenotypic space of aphasia?
This question arose as a result of the different test batteries employed to compare aphasia across the thesis. In Chapter 2, PSA and PPA were compared on the same extensive test battery which assessed a variety of linguistic domains including phonological discrimination, repetition (of words, non-words, and sentences), naming, semantic memory, sentence comprehension, and production of connected speech. In Chapters 3 and 4, PSA and PPA/movement disorders were assessed using the Mini-linguistic State Examination (MLSE) which evaluated the same linguistic domains, plus repetition of syllables, oral reading (words and non-words), and writing. However, the shared battery from Chapter 2 also assessed a variety of ‘non-linguistic’ domains, namely memory, attention/executive function, visuospatial processing, and praxis.

By measuring these non-language domains, the PCA method showed that there was important variation in these domains in PSA and PPA – i.e., variance in visuo-executive function. An executive component was not extracted by PCA on the MLSE battery or Matched battery from Chapter 3 since neither battery contains measures of executive function. Whilst non-language impairments are not part of the diagnostic decision tree for PPA or PSA, it is clear that they represent a significant aspect of the phenotypic space of both forms of aphasia (Butts et al., 2015, Murray, 2012). Previous studies which have included explicit measures of attention and executive function, have found an executive function dimension extracted by PCA (Butler et al., 2014, Halai et al., 2017, Lacey et al., 2017). In fact, when non-verbal executive function is assessed in more detail, more nuanced dimensions reflecting different aspects of executive function can be extracted from variance in this cohort (Schumacher et al., 2019). This is important because executive deficits have been shown to underlie impairments in language tasks, such as semantic tasks (Thompson et al., 2018), and also syntactic processing (Tan and Martin, 2018). Furthermore, in Chapter 5, the inclusion of tests of visual function revealed variation in visual function even in the cases of typical AD. This is in line with previous studies of typical AD which have found important variation in domains like language, vision and behaviour, when these non-memory related symptoms are assessed (Lambon Ralph et al., 2003, Peter et al., 2014).

The significance of these findings is that they show the importance of measuring more than just the expected dimensions of variance in a neurological condition. However, since the intention of studying different neurological conditions is often to differentially diagnose, the results of the thesis suggest that measuring non-target domains might aid this. For example, activities of daily living (Jang et al., 2012), auditory processing (Grube et al., 2016), visuospatial processing (Watson et al., 2018) and empathy (Hazeldon et al., 2017) have all been shown to be altered in different forms of PPA.
This is also important in progressive conditions because as the degenerative disease progresses to more distant regions of the brain, eventually all cognitive faculties become impaired (Leyton et al., 2016a). Although there is a focus on assessing people in the early stages of the disease (e.g., for differential diagnosis, or clinical trials for disease-modifying therapies), understanding progression and prognosis also requires assessing late-stage symptoms. In these later stages it will be important to assess beyond the expected dimensions of impairment.

However, assessment of extraneous cognitive symptoms in aphasics is unrealistic in the clinical setting. Thus, the MLSE – as a novel clinical assessment tool for PPA – is being developed to assess only the linguistic domains proposed in the diagnostic consensus recommendation to be important for differential diagnosis of the subtypes. It is possible that in the clinic, tests of other cognitive functions could be assessed with further testing in cases who are not easy to diagnose.

Limitations of the approaches applied in the thesis

Principal component analysis was chosen to uncover the latent structure of variance in the patient cohorts of the research in this thesis because it is a data-driven analysis which does not attempt to find categories/clusters within the data. However, there are a number of limitations which must be taken into consideration regarding the PCA technique. One of the drawbacks of exploratory factor analysis techniques, including PCA, is that the resultant solution is necessarily limited by the information entered into the PCA. To explain this further, one can use the example of the results of Chapters 2 and 4; in Chapter 2, tests measuring visuospatial processing, and attention and executive function were included in the PCA, whereas, in Chapter 4, the MLSE did not assess any non-language capabilities. Therefore, the PCA in Chapter 4 could never replicate the visuo-executive component found to be important in describing the multidimensional space of PSA and PPA in Chapter 2. This limitation of PCA was further demonstrated by Halai et al. (2018b) who used an extensive test battery of naming errors and PCA to uncover a 5-factor solution encompassing semantic, phonological, dysfluency, circumlocution, and omission errors. The authors replicated the different number and composition of components found in previous studies by reducing their test battery; for example, when removing their tests of non-language function, the authors were able to replicate the results of Mirman et al. (2015a, 2015b). Thus, it is important to assess beyond the expected domains of impairment in any patient cohort, as discussed above.

A related limitation of PCA occurs when the data entered into PCA are dominated by data from assessments of certain cognitive domains (Kim, 2008). For example, over-representation of expressive phonological tests (e.g., repetition) necessarily increases the relative contribution of variance from this cognitive domain, and therefore the resultant solution reflects this. Similarly, other cognitive domains will be under-represented, and therefore contribute less variance to the solution. This is why it is vital to create a balanced
dataset (in terms of the number of tests measuring different cognitive-linguistic domains) as far as possible.

Another consideration which is important for PCA is the selection of the number of components in the final solution (Coste et al., 2005). The reason this is important is that over- or under-extracting factors can heavily obscure the results (Wood et al., 1996). Many studies using PCA employ the traditional Kaiser criterion (Kaiser, 1958) of eigenvalues >1. This equates to retaining components only if they account for more variance than a single input variable. We used a k-fold cross-validation procedure to determine the optimum number of components to retain. This approach has been shown to avoid over-extracting factors, which can occur with other methods (Coste et al., 2005).

In this thesis we have used PCA to extract components which we believe reflect real cognitive processes which are supported by specific brain regions or networks. We have applied varimax rotation in order to aid cognitive interpretability of the extracted components, to attempt to relate them to the potential underlying cognitive processes which we believe could be supporting performance on the varied cognitive tests we have administered. A key aspect of this approach is that we necessarily assume that the cognitively interpreted components will have neural correlates. Thus, the logical and necessary next step when using this PCA approach is to relate the components to the brain to independently validate their interpretations. Without this form of validation, the extracted components may be artefacts of the PCA (given the limitations discussed above) and may not reflect real cognitive abilities which arise due to activity of real neural networks. We have attempted to independently validate the components in Chapter 4 by relating them to the brain using VBCM. Chapter 4 made use of VBCM to uncover the neural correlates of the behavioural components extracted by PCA. This is a form of voxel-based morphometry (Tyler et al., 2005) which treats neural integrity as a continuous variable, as opposed to voxel-based lesion-symptom mapping (Bates et al., 2003) which binarizes the neural integrity of each voxel as spared or damaged. This approach was important as it made it possible to relate scores on the continuous PCA dimensions to continuous neural integrity values. Both VBCM and voxel-based lesion-symptom mapping are mass univariate methods in which statistical testing is done at the level of each voxel to see if damage in that voxel is associated with the behaviour in question (Wilson and Hula, 2019). There are a number of limitations associated with these approaches which need to be taken into consideration for the results of Chapter 4.

Firstly, there are a number of technical concerns regarding systematic differences between healthy and diseased brains (Mechelli et al., 2005). There may be group-level differences which create systematic confounds (such as patients moving more in the scanner) which can be problematic if not controlled for. Alternatively, atypical brains (due to stroke lesions or neurodegeneration, for example) may not match the templates built into the VBCM pre-processing pipeline, which could result in a systematic issue with misregistration or
misalignment in the patient group. To mitigate these issues, it is important to include covariates such as lesion volume, scanner model, etc. so that potential systematic differences between groups can be recognised and accounted for.

Secondly, there are a number of reasons why VBCM is not suitable for detecting network-level associations between behavioural scores and neural abnormality. One reason is that because VBCM reveals the unique neuroanatomical correlates associated with each variable of interest, this approach inherently misses neural correlates which are related to more than one variable of interest. Another reason is that voxel-based morphometry methods which employ smoothing at the level of each voxel are inherently biased towards detecting effects which are spatially localised (Davatzikos, 2004). Furthermore, as a mass univariate method, VBCM is looking for associations at each voxel individually and so fails to take into account spatial contingencies between neighbouring voxels (Wilson and Hula, 2019). To account for this in Chapter 4, we added ‘lesion volume’ as a covariate, however this does not totally remove this confound. Another reason the VBCM is not suitable for detecting associations with networks is that mass univariate approaches like VBCM uncover regions where reduced neural integrity is associated with a variable of interest, yet do not consider the effects of disconnection resulting from damage in these regions. This could mean that the observed association between a behaviour and a region is actually being driven by damage in another region which has become disconnected from the observed region. Multivariate methods, such as support vector machines (Murley et al., 2019), are needed to take into account contingencies between voxels and also interactions between multiple brain regions/networks.

One of the novel contributions of Chapter 4 was applying VBCM to PSA and PPA in the same neuroimaging analysis, having assessed neural integrity in the same way for the stroke-lesion and neurodegenerative brains. To achieve this, we used an automated lesion identification procedure which was developed and validated for stroke lesions. The results of Chapter 4 suggest that this approach was successful in identifying regions of significant atrophy in the neurodegenerative brains. However, due to the above limitations of VBCM and the exploratory nature of the study in Chapter 4, further research is needed to validate the findings. For example, the automated lesion identification procedure (designed for stroke lesions) could be compared to the results gained by quantifying neural integrity with cortical thickness measurements (more extensively used for brains with neurodegeneration) (Li et al., 2015). Furthermore, multivariate measures could be used to model (a) the effects of a stroke lesion which results in damage to all regions in the occluded vascular territory, including long-range white matter fibres passing through and linking distant brain regions (Wilmskoetter et al., 2019), and (b) the effects of neurodegenerative pathology spreading through particular layers of the cortex or within networks (Seeley et al., 2009).
Compatibility of the PCA approach with single-case and case-series approaches

Two traditional methodological approaches in cognitive neuropsychology are single-case studies and case-series studies. Single-case studies make use of extensive test batteries to probe deficits in one case (or a small number of cases) in great detail, whilst case-series make use of larger, more representative samples and thus are able to account for individual differences. According to Schwartz and Dell (2010), the goal of case-series studies is to use patterns of co-variance to address questions about complex cognitive mechanisms, and vital to this aim is the inclusion of a heterogeneous sample recruited with lenient selection criteria. Furthermore, the result of a case-series approach should be pertinent at the level of individual cases (i.e., individuals’ data should be at least recoverable) (Rapp, 2011). These features mean that case-series approaches can avoid the pitfalls of averaging across individuals, which is one of the main reasons why many researchers prefer single-case methodology (e.g., Caramazza, 1986).

The PCA approach used in this thesis requires highly detailed assessment of diverse cognitive abilities in a large cohort of heterogeneous patients. As such, this approach shares features with both single-case studies and case-series studies. Firstly, as with single-case studies, this PCA approach benefits from in-depth testing to gain a picture of the full spectrum of behavioural performance (both spared and impaired). Secondly, as with case-series studies, this PCA approach capitalises on heterogeneity between individuals when lenient selection criteria are used for the large cohorts studied. The PCA approach in this thesis is also similar to case-series studies in that the result (visualising factor scores in the multidimensional space) is still at the level of individual cases (i.e., not group averages).

There is on-going debate as to the relative importance of contributions from single-case or case-series approaches for cognitive neuropsychology (Rapp, 2011, Lambon Ralph et al., 2011). Choosing between these approaches is non-trivial, since there is often a trade-off between having a more representative sample vs. more in-depth testing (Rice et al., 2020). On the one hand, researchers may prioritise a larger, heterogeneous cohort tested on a smaller battery (potentially only assessing the target domain and therefore not assessing the full phenotypic picture). Alternatively, researchers may prioritise a more detailed test battery, which, due to its exhaustive length, means fewer cases are studied. In this latter example, researchers may use strict recruitment criteria (e.g., based on diagnostic subtype) to select a smaller, homogeneous sample (or single case) to administer this extensive battery to.

The PCA approach used in this thesis can inform researchers who are faced with this trade-off, to make the selection of tests or cases more efficient. For example, having uncovered the structure of variance in a particular patient cohort using PCA, the result can inform (1) recruitment of representative samples from all regions of the phenotypic space, and (2) efficient selection of in-depth tests to assess all dimensions of the phenotypic space of a disorder in a small number of tests.
Without administering all the tests that went into the PCA solution, the exact position of newly recruited cases in the extracted multidimensional space cannot be precisely quantified (perfectly comparable factor scores cannot be generated without scores on all the original tests). However, the behavioural profile of newly recruited cases can be qualitatively compared to existing cases exemplifying different regions of the multidimensional space. For example, from Chapter 5 it was clear that there are cases of posterior cortical atrophy who occupy all corners of the phenotypic space as defined by visuospatial and visuoperceptual impairments. This means that to sample the full phenotypic space with a representative cohort, recruited cases should display (a) spared visuospatial processing but impaired visuoperceptual processing, and vice versa, (b) mild impairments in both domains, and (c) severe impairments in both domains. By using the PCA solution to inform recruitment of more representative samples, a smaller cohort can be tested whilst ensuring that the full phenotypic space is represented.

In terms of selecting the most informative tests, the tests with the strongest factor loadings onto the extracted components contribute most of the variance captured by those components. This means that performance on the test with the strongest loading on the phonology component, for example, provide the closest approximation of performance on the phonology component itself. Thus, having uncovered the dimensions of variance important for a particular disorder using an in-depth and extensive test battery, the PCA solution can inform the selection of a smaller battery which can approximate the established multidimensional space. Halai et al. (in prep.) used PCA to create a reduced test battery for fair comparison with the Comprehensive Aphasia Test (CAT) (Swinburn et al., 2004). The purpose of their study was to compare the underlying structure of variance captured by the CAT and the structure captured by a more in-depth test battery. However, to match the brevity of the CAT, Halai et al. used PCA to create a reduced battery from the extensive starting test battery. To accomplish this, they ran PCA on the data from the extensive test battery, then selected only the two tests with the highest loadings on each extracted component. These tests contributed the most variance to each component, and thus, using data from only these reduced number of tests gave an approximation of the underlying structure, in less time than required to administer the full extensive test battery.

Aside from informing selection of representative samples and informative tests in new studies, the PCA approach is naturally compatible with follow-up single-case investigations, because individual participants’ data are the focus of the PCA approach (i.e., plotting individuals’ factor scores in the multidimensional space). For example, the behavioural profiles of patients occupying overlapping regions of the multidimensional space can be further characterised and explored, regardless of their diagnostic label. For example, in Chapter 5 we showed that some typical AD cases displayed relatively spared general cognitive status coupled with impaired performance on tests of visual processing. A single-case investigation approach could have been applied to explore what was driving visual
processing impairment in one of these cases. This approach was used by Lambon Ralph et al. (2003) to further characterise the profiles of individuals who showed impairment in a non-memory domain which could not be explained by their general level of severity as detected by PCA. Additionally, individual patients with interesting or unexpected behavioural profiles can be visualised in the context of the full phenotypic spectrum revealed for the group. This approach has recently been used to investigate potential category-selective deficits in posterior cerebral artery stroke; Rice et al. (2020) combined the data-driven PCA approach with case-by-case quantification of deficits using composite scores. This allowed them to show that single cases displaying category-selective deficits were the exception, not the rule, in their large, heterogeneous cohort.

Overall, the PCA approach shares features with single-case and case-series approaches, and has the potential to inform studies which are faced with the trade-off between more representative samples and more in-depth testing. Therefore, the PCA approach can be used to enhance the fruitfulness of single-case and case-series research.

**Clinical applications**

The PCA framework employed in this thesis is useful for understanding patient variance in a research setting wherein a large cohort can be tested on an extensive test battery, but this is not always the case in a clinical setting, such as a clinical trial. The PCA approach we have used is not aimed at replacing differential diagnosis using clinical assessment tools on an individual level. However, we believe that this approach can have a number of clinical applications, both immediately in the short-term use of the approach to inform upcoming studies, and longer-term applications to inform the continual evolution in the development and use of diagnostic classifications. These potential clinical applications are outlined below.

The PCA approach in this thesis has a number of clinical applications which can be capitalised on in the short-term. Firstly, since PCA can reveal graded variation between proposed diagnostic subtypes, this approach can be useful to clinicians as a way of visualising what they experience in the clinic, namely highly heterogeneous and overlapping presentations of cognitive/linguistic/visual impairments on a case-by-case basis which are often hard to differentially diagnose. In that regard, this PCA approach may provide empirical evidence to reflect the intuitive understanding that clinicians possess already (rather than providing another attempt to carve variation into diagnostic categories which may or may not reflect clinical experience).

Secondly, the PCA approach, when applied transdiagnostically to multiple aetiologies (e.g., stroke aphasia and PPA), can reveal where labels and terminology are used, by clinicians and researchers, with equivalency or not. This was highlighted in Chapter 2 with regards to the motor speech production component, which captured aspects of connected speech production. The PSA subtypes considered to be ‘fluent’ within the spectrum of PSA tended
to have worse performance on this component than the ‘progressive non-fluent aphasia’ cases. This difference in ‘fluency’ has been demonstrated before (Patterson et al., 2006a) but the PCA approach enabled the motor speech production performance of individuals to be mapped out in the context of the transdiagnostic, shared multidimensional space, to show the variability within and between non-fluent subtypes of PSA and PPA. There may be other labels or terms which are applied across different aetiologies of cognitive/linguistic disorders, which may not mean the same thing within the context of each disorder. The transdiagnostic PCA approach provides an appropriate platform for investigating these differences in the use of clinical labels and terms, without grouping patients based on their diagnostic labels.

Thirdly, the PCA approach can benefit future clinical studies and clinical trials in two ways: (1) guiding the selection of the most informative tests and assessments to capture the core, underlying dimensions of variance as uncovered by PCA, and (2) informing the recruitment of cases to ensure that all regions of the previously-established phenotypic space of a condition are sampled equally (i.e., to avoid over-sampling of particular sub-regions of the true multidimensional phenotypic space). These benefits have been discussed above in the context of informing the trade-off between more representative samples and more in-depth testing (Rice et al., 2020), which is especially relevant for clinical trials.

Multiple factors in a clinical trial could put pressure on this trade-off, particularly on selecting the most informative tests for a brief test battery, such as patient stamina or time, training demands on research staff, or having multi-centre protocols each with their own outcome measures (Lageman et al., 2010). Selecting a homogeneous sample using strict exclusion criteria is often a priority in clinical trials in order to obtain more precise estimates of effect sizes, however this can reduce the generalisability of the results beyond the studied sample (Britton et al., 1999). Therefore, selecting a representative cohort which samples the full phenotypic space of a disorder is important, and PCA can be used to achieve this in two ways. Firstly, PCA can be used to ensure that all regions of the multidimensional space are being sampled equally, by placing individuals in the context of the full phenotypic space (as in Rice et al., 2020). Secondly, if a homogeneous sample is required, PCA can be used to select cases which have overlapping multidimensional profiles (i.e., occupy the same region of the multidimensional space) rather than using diagnostic subtype labels for inclusion criteria. This means that, irrespective of their clinical diagnosis, patients with relatively homogeneous behavioural symptoms could be selected based on occupying a shared region of the multidimensional space (thereby sharing symptomatology across the cognitive systems captured by the dimensions). Such a transdiagnostic approach has the potential to inform clinical trials for disease modifying treatments because it can isolate common symptom profiles across different diagnostic categories with a shared underlying pathology (Murley et al., 2019, Passamonti et al., 2018). Thus, the PCA approach employed in the thesis has several potential clinical applications which could be capitalised on in the short-term, by upcoming clinical studies and clinical trials.
We foresee that the PCA approach in the thesis could also have longer-term clinical applications through informing the continual evolution in the development and use of diagnostic classification systems. Although we argue that the PCA results of the thesis highlight where categorical diagnostic classifications are currently less useful, we are not proposing that diagnostic categories be abandoned entirely. Instead, it may be the case that the PCA approach, which accounts for graded variation, can improve the transdiagnostic investigation of relationships between behavioural profiles and underlying pathological mechanisms driving impairments. For example, PCA provides an ideal framework for relating graded behavioural variation to other clinically relevant markers of disease and disorder, such as integrity of neural networks, or build-up of pathology. Murley et al. (2019) combined PCA and support vector machine multivariate analyses to find shared symptom dimensions and co-varying regions of brain atrophy, which did not reproduce the categorical differences between the included diagnostic subtypes. However, this approach constitutes a framework for investigating network-level brain atrophy caused by fronto-temporal lobar degeneration regardless of subtype diagnosis, thus improving our understanding of the relationship between this form of neurodegeneration and atrophy of specific neural networks. Further research could build on this to refine differential diagnosis of subtypes of fronto-temporal lobar degeneration based on proportion or distribution of damage to these networks. In this regard, we believe the PCA approach has the potential to inform refinement and development of differential diagnosis in the long-term through further research.

**Directions for future research**

The transdiagnostic, multidimensional approach applied in this thesis highlights that considering patient variance along continuous dimensions instead of categorical systems has benefits for (a) comparing directly the full phenotypic space of different disorders, and (b) relating principal behavioural dimensions to unique neural correlates of structural brain abnormalities.

A strength of the studies in Chapters 2 and 4 is the inclusion of heterogeneous cohorts of aphasia from multiple aetiologies, which enabled shared dimensions of variance and neural correlates to be examined transdiagnostically. Chapter 2 made use of an extensive test battery measuring multiple aspects of language and cognition yet lacked neuroimaging evidence to validate the extracted principal dimensions of variance found to be in common for PSA and PPA. Chapter 4 made use of a brief clinical assessment tool with fewer subtests and items, thereby assessing a narrower range of linguistic abilities (and no assessment of non-linguistic impairments). However, Chapter 4 did benefit from structural neuroimaging analysis to relate the extracted principal dimensions to unique neural correlates.
Two avenues of research could build upon these results. Firstly, future research could explore the syntax component further by including more tests of grammatical processing (e.g., tests from the Northwestern Assessment of Verbs and Sentences (Thompson, 2011)), and more tests of motoric speech production (e.g., measures of apraxia of speech such as the Apraxia of Speech Rating Scale (Strand et al., 2014)). Including more variation from tests in these domains could result in the syntax component splitting into separable components reflecting grammatical processing and verbal speech output. The neural correlates of the latter could then be compared with the motor speech production component of the primary systems hypothesis (Woollams et al., 2018). This line of research would also be highly relevant to researchers who propose that there are separable apraxic and agrammatic presentations of PNFA (Tetzloff et al., 2019, Josephs et al., 2012), because it could reveal the nature of variation along potential apraxic or agrammatic dimensions, in PNFA and also in non-fluent PSA as a comparison.

Secondly, an outstanding question from the results of Chapter 4 is whether an individual's position in the aphasic multidimensional space can be related to the degree of neural abnormality in expected brain networks. Several studies have shown that presentations of neurodegenerative disease can be linked to the integrity of dissociable neural networks. Warren et al. (2012) proposed that variants of Alzheimer’s disease may be characterised by the pattern of relative integrity of different nodes in the default mode network. Furthermore, Seeley et al. (2009) found dissociable patterns of atrophy associated with different neurodegenerative disease syndromes. Murley et al. (2019) found co-varying relationships between shared symptom dimensions and regions of brain atrophy in fronto-temporal lobar degeneration syndromes. Furthermore, the spatial distribution of damage to the language network has been related to presentations of PPA (Gorno-Tempini et al., 2004) and PSA (Tippett and Hillis, 2016). However, most of these studies have related measures of network integrity to patients’ symptom profiles grouped by their diagnostic subtype (with the exception of Murley et al. (2019)). Therefore, future research could combine PCA with multivariate imaging methods to explore relationships between integrity of networks and graded, multidimensional profiles of impairment.

One application of this approach would be to explain an individual’s profile of spared/impaired domains based on their position in the MDS and their pattern of neural abnormality. For example, Halai et al. (2017) showed that whether an individual PSA patient showed spared or impaired fluency could be related to whether their lesion overlapped the VBCM cluster for fluency. Thus, future research could use PCA and multivariate analysis methods to establish a network of brain regions supporting the core underlying dimensions of variance in aphasia, then relate an individual’s position in the multidimensional space to the degree of overlap of their pattern of neural abnormality with the established networks.
The study in Chapter 5 highlighted a spectrum of visual impairment with graded differences between typical AD and posterior cortical atrophy cases. Graded differences in visuoperceptual and visuospatial impairments were also found within posterior cortical atrophy, providing evidence against the existence of categorical perceptual/spatial subtypes. Further research is required to independently validate the dimensions extracted by PCA, for example through replication with other cohorts and test batteries, and through neuroimaging analysis to relate the components to neural correlates. For example, the general cognitive status components require further investigation because they contain loadings from all the ‘non-visual’ assessments, and therefore capture variation from heterogeneous cognitive domains such as memory, attention, and language.

Future research could also use PCA and neuroimaging methods to explore the relationship between variants of AD and their patterns of atrophy in two ways. Firstly, as discussed above, the PCA approach is naturally compatible with follow-up single-case investigations. Therefore, a follow-up study from Chapter 5 which included neuroimaging data could investigate the behavioural and neural profiles of the cases of typical AD who showed relatively more impaired visual processing ability than general cognitive impairment (i.e., those who overlapped with mild posterior cortical atrophy in the multidimensional space). This approach has been demonstrated by Rice et al. (2020) who investigated category-specific deficits in individual cases of posterior cerebral artery stroke and related this to their lesion locations. These authors showed that laterality of the posterior cerebral artery stroke was related to the position of individuals in the two-dimensional space defined by relative deficits in face or word processing, as extracted by PCA. Using this approach with variants of typical and atypical AD could elucidate the relationship between degree of overlap on visual processing impairments and degree of overlap in the pattern of temporo-occipital or parieto-occipital atrophy.

Secondly, future research could build on the transdiagnostic PCA approach employed in Chapter 5 by including other variants of AD and including multivariate measures of network-level neural integrity. This could be used to explore the neural network paradigm of AD, in which phenotypic variation in Alzheimer’s disease may be related to the pattern of relative integrity of different nodes in the default mode network (Warren et al., 2012). Future research could therefore aim to (a) establish the full phenotypic space of AD by applying PCA to a heterogeneous cohort of AD presentations assessed on an extensive neuropsychological battery, and (b) relate regions of the extracted multidimensional space (i.e., different behavioural profiles) to the pattern of atrophy in distinct nodes of the default mode network. This research could inform further studies relating underlying pathological and genetic mechanisms to phenotypic variation in AD, with the goal of informing research into disease modifying treatments.
Conclusion

In conclusion, the research in this thesis has shown that behavioural variance in patients with neurological and neurodegenerative disease conforms to an underlying continuous structure which can be extracted with principal component analysis. Variation in language impairments in aphasia can be captured with three underlying language-cognitive dimensions, reflecting phonology, semantics and connected speech/syntax. Behavioural impairments in typical and atypical variants of Alzheimer’s disease can be captured by core underlying dimensions reflecting general cognitive status and visual processing. Conceptualising behavioural variance in these conditions along the continuous dimensions can account for mixed and heterogeneous presentations within subtypes, and also highlight blurred boundaries between subtypes.
References


DÉJERINE, J. J. & SÉRIEUX, P. 1897. *Un cas de surdité verbale pure, terminée par aphasia sensorielle, suivi d’autopsie*, publisher not identified.


HALAI, A. D., DE DIOS PEREZ, B., STEFANIAK, J. & LAMBON RALPH, M. A. in prep. Comparing the underlying structure of short and long aphasia batteries to assess deficits in chronic post-stroke aphasia.


syndromes associated with posterior atrophy early age at onset AD spectrum. *Neurology*, 73, 1571-1578.


REY, A. 1941. L’examen psychologique dans les cas d’encéphalopathie traumatique. (Les problems.). Archives de psychologie.


